BERING SEA TANNER CRAB STOCK ASSESSMENT REVIEW FOR THE CENTER OF INDEPENDENT EXPERTS

Dr Cathy Dichmont

Cathy Dichmont Consulting
cathydichmont@gmail.com

Review of the Bering Sea tanner crab assessment

1 CONTENTS

1.2 LIST OF FIGURES

1.3 LIST OF TABLES

2 Executive Summary

The review workshop for the tanner crab assessment took place in Seattle on 31 July to 3 August 2017 with review panel members Drs Dichmont, Hall and Nielsen, the tanner crab assessor, Dr William Stockhausen, and other scientists involved in the stock assessment, data collection and management. The review documents have been placed on a Google Drive folder and several model test outputs for the Review. Several very insightful presentations were provided during the review, with very open and free flowing discussion. These greatly contributed to the reviewers' knowledge base.

During the review, several further output diagnostics were requested as well as further information on model time blocks and priors. A few sensitivity tests were also requested and provided. The interaction between the review team and the review attendees was extremely informative and helpful. Dr Stockhausen should be thanked for his enormous effort in providing information and further runs, often after hours. This is greatly appreciated.

The TCSAM02 framework is supported. It is flexible and able to undertake additional tests without changing the code. The input files (once understood) are reasonably intuitive. The documentation provided on TCSAM02 is reasonably complete and clear. However, there is some confusion as to what the current model settings are in the documentation. Some of this confusion occurs when the assessor attempts to describe both the framework and current assessment together. Separating these two components should help the reader. The mathematics in the model, in the main, is appropriate, especially the recent inclusion of GMACS components.

However, the model appears to be too constrained and uses too many different approaches for this process. This is not supported. Solutions are not unique and are overly dependent on input settings. A suggested improvement is provided.

The projections are undertaken appropriately. Stochastic versions and choosing a more recent time period as sensitivity tests are recommended.

The tanner crab assessment is data rich, although complicated in that tanner crab are caught as a bycatch in several fisheries as well as the directed fishery. Survey data are extensive and cover most of the spatial range of the species, especially after 1987. Total catch, discard and size frequencies are obtained from the fleets either through at-sea observers or dockside. Observer coverage is good. Classic statistical (as opposed to model based) aggregation techniques are used to the highest standard. On the whole, there do not seem to be major issues with regard to the quality of the data and they are generally appropriately used. Despite this statement, some improvements can be made, with regard to simulation testing and constructing a different approach to the model estimation process.

Potential uncertainty in the foreign catch data was tested and shown to not have large impacts on the post-1982 model outputs. This uncertainty is therefore not a priority.

The Crab Plan Team (CPT) has produced a very rigorous set of guidelines (prepared by Punt and Kinzey) of which diagnostics and tests should be provided. Those provided are supported. Extensive diagnostics are provided as a norm when considering the 2016 assessment report. Undertaking MCMC is recommended, but only when the model is less restricted. Jitter runs were not fully investigated. Results during the review showed that this is important. Profile likelihoods were attempted, but an ADMB bug made this impossible during the review. These should urgently be investigated given the model test results. Different additional retrospective tests are suggested.

Including the side-by-side survey data in the assessment is supported. The approach suggested by Buckhausen as a first cut is also supported. Estimating within site variability for the nephrops surveys is recommended. A few alternative approaches to the data inclusion are also suggested. Continuing this data source is recommended, given the short-term utility of the data in terms of selectivity and also relative catchability.

While the model is too restricted, there are clear conflicts with the data, given the correlations and bounds being hit. A series of steps to move forward are suggested. These include removing any bounds or priors that could be solved through transformation, calculating the effective sample size and using a sampling design to test the impact of different sample sets. However, major structural changes are not recommended without undertaking a simulation study, which is highly recommended. In-depth suggestions regarding model structure are only really feasible when the current model and its variants are tested using simulated data.

The present research underway on growth, movement and bitter crab syndrome are supported as very important in the short and medium term. A first step of undertaking a model approach (external to the assessment) of the spatio-temporal changes in the biology and fleet of tanner crabs is recommended. A suggested medium-term priority is to undertake end-end whole of systems ecosystem modelling. The region is well studied with extensive data, and yet mortality (predation) and oceanography play an important role in the spatio-temporal distribution of the species. Oceanographic modeling is also suggested, including ones that relate to larval distribution. An example of such a model is provided. Although economic information currently seems less important, there are some suggestions regarding comparison of on-board observer view of size and shell condition animals that are likely to be discarded and the dockside observer data.

A major recommendation is to undertake simulation testing of the model. A series of points are provided that highlight some of the options that should be tested. This is the highest priority recommendation.

4 BACKGROUND

Tanner crab (*Chionoecetes bairdi*) are caught in the Eastern Bering Sea as a directed tanner crab pot fishery and as bycatch in snow crab (SCF), Bristol Bay red king crab fishery (RKF) and the groundfish trawl bycatch fishery (GTF). A key biological characteristic of tanner crabs is that they reach a terminal molt at maturity. Annual crab and finfish directed trawl surveys have been undertaken for several decades and are key source of information. At sea and dockside observers also collect numerous forms of data from the fishing vessels and processors.

The tanner crab assessment is an integrated size-based model that is disaggregated by sex, shell condition and maturity. One-year projections are undertaken to calculate the OFL and ABC. Since there is no stock-recruitment relationship for tanner crabs, they fall within the Tier 3 category for the Crab Plan Team system. The Tier 3 control rule uses $B_{35\%}$ and $F_{35\%}$ as proxies for BMSY and FMSY respectively.

A series of assessments are provided – what is referred to as the assessment framework. This framework has been developed and expanded over the past few years and provides a flexible environment where options can relatively easily be changed without changing the code. R scripts are also used to generate diagnostics.

Recent side-by-side surveys have been undertaken in the EBS, where a vessel with nephrops gear trawls alongside the NMFS trawl survey vessel. These data are not yet included in the tanner crab assessment. Extensive research on growth and mortality of crab particular to the EBS is underway, enhancing the data set and replacing data from other regions.

The fishery assessment falls under the remit of the Crab Plan Team (CPT) that has over the years provided direction on which changes to the model are high priority and what tests are to be undertaken. The assessment is still under-going change and it is important to note that this review was during an assessment year and not the final assessment for OFL setting in 2017.

6 DESCRIPTION OF THE INDIVIDUAL REVIEWER'S ROLE IN THE REVIEW ACTIVITIES

The review workshop for the tanner crab assessment took place in Seattle on 31 July to 3 August 2017. In attendance were review panel members Drs Dichmont, Hall, and Nielsen, the tanner crab assessor, Dr William Stockhausen, and other scientists involved in the stock assessment, data collection and management. Dr Stockhausen provided the documents on a Google Drive folder and several model test outputs for the Review (Appendix 1). The Statement of Work provided to the review panel is provided in Appendix 2. The chair was Dr Dorn. Participants of the workshop are provided in Appendix 3 and the Agenda in Appendix 4.

Several very insightful presentations were provided during the review, with very open and free flowing discussion. These greatly contributed to our knowledge base.

During the review, several further output diagnostics were requested as well as further information on model time blocks and priors. A few sensitivity tests were also requested and provided. The interaction between the review team and the review attendees was extremely informative and helpful. Dr Stockhausen should be thanked for his enormous effort at providing information and further runs, often after hours. This is greatly appreciated.

There was some discussion as to which model is the current model and this was decided during the workshop.

No panel report was required. An individual reviewer's report was provided addressing each Terms of Reference (ToR), being:

1. Statements assessing the strengths and weaknesses of the current Tanner crab stock assessment model with regard to population dynamics, fishery and survey components, likelihood components, and model evaluation.
2. Statements assessing the strengths and weaknesses of the current Tanner crab stock projection model, with regard to methodology.
3. A review of the fishery dependent and independent data inputs to the stock assessment with regard to quality of information and appropriateness to the assessment.
4. Recommendations for alternative approaches to evaluate model convergence and compare multiple models.
5. Recommendations for integrating BSFRF surveys into the assessment.
6. Recommendations for alternative assessment/projection model configurations.
7. Recommendations for research that would reduce the uncertainty associated with key parameters assumed or estimated in the assessment.
8. Suggested priorities for future improvements to the stock assessment/projection model.

7 SUMMARY OF FINDINGS FOR EACH TOR IN WHICH THE WEAKNESSES AND STRENGTHS ARE DESCRIBED

7.1 TOR 1: STATEMENTS ASSESSING THE STRENGTHS AND WEAKNESSES OF THE CURRENT TANNER CRAB STOCK ASSESSMENT MODEL WITH REGARD TO POPULATION DYNAMICS, FISHERY AND SURVEY COMPONENTS, LIKELIHOOD COMPONENTS, AND MODEL EVALUATION.

7.1.1 Model framework and process

A series of reports describing model version and numerous sensitivity tests undertaken in 2016 and 2017 were provided to the review team. Model code and results were also provided in three sub-directories. All these have similar basic model dynamics, being disaggregated by sex, shell state, size and maturity status. A key component of the assessment model framework is the assumption that animals undergo their terminal molt to maturity which is an essential element of tanner crab biology.

The documents describe the assessment model run in 2016 called TCSAM2013 and the transition to the newer model used for 2017 called TCSAM02. A very rigorous process is described in the transition from the TCSAM2013 to TCSAM02 ensuring that the results are repeatable and as bug-free as possible. The key to this is that TCSAM2013 and TCSAM02 were updated so that they could be set up as similarly as possible to test whether they provided essentially the same results. This process was successful.

In the main, the effect of these transitions from old to new is generally well described and tested. This provides great confidence that the latest model is as bug-free as possible and can repeat past analyses. Under the direction of the Crab Plan Team (CPT), there is a clear and rigorous stepwise process starting with repeating the old and new model runs to get the same results, and adding the new components to TCSAM02 (mostly) one at a time to understand the impact of each change. This conforms to world's best practice.

The ability to undertake numerous sensitivity tests has been greatly enhanced by changes started in TCSAM2013 and mostly completed in the TCSAM02 code structure and using R code to automatically produce the required graphs and documents. This means the architecture of TCSAM02 is much improved from that of TCSAM2013. This would have required a large effort to make the model as flexible as possible in the code and removing any components that are hard wired. This is a *major* improvement from TCSAM2013 and the assessor is congratulated on successfully undertaking this transition.

Furthermore, the assessor is clearly also adopting key components of the GMACS framework (https://github.com/seacode/gmacs/wiki) where the intent is ultimately to adopt GMACS. However, at this stage, GMACS is not suitable for use with animals that undertake a terminal molt such as tanner crabs. As a result, importing components from GMACS into TCSAM02 is supported. This avoids re-invention and provides more consistency between crab assessments.

Recommendation 1. **Given the successful transition from a more flexible model framework, it would be a retrograde step if GMACS does not adopt a similar approach. As a result, it is not recommended that another framework such as GMACS be adopted unless this is similarly flexible.**

Although there is a large overlap between the two models, there are also substantial changes in the settings and structure of TCSAM02 compared to TCSAM2013. The most obvious of which is: a) to remove hard wired components from the model into files that are read into the code, b) allowing for more flexibility in the number of time blocks defined for model parameters, as well as the number of fisheries and surveys, c) providing alternative selectivity functions and prior functions, d) including the GMACS growth model, e) calculating the OFL directly in the model, and f) including the GMACS approach to modelling fishing mortality. These changes and updates allow for a much more flexible model that is much easier to update and change.

The general population dynamics described within the framework is supported and is similar in most part to that undertaken in many size-based assessments. Clearly, not all options within the framework are equal, some are included for legacy reasons and should not be selected and others are updates that provide better model fit characteristics.

The model framework is of a high standard.

7.1.2 Model description

Several documented descriptions of both TCSAM2013 and TCSAM02 model were provided. The frameworks of TCSAM2013 and TCSAM02 are well described, providing mathematical descriptions of functional form options in the code. The 2016 base settings for TCSAM2013 are also provided (Appendix C in Stockhausen, 2017).

However, the same is not true for TCSAM02. The framework is reasonably well described. The current (base) model TCSAM02a settings are not well described. The documentation often confuses describing the model with the options available in the framework. This caused some confusion as to what the exact settings for the base case model were and for the model tests provided. This, in turn meant that the input files provided had to be read to fully understand *exactly* what the settings were. This may be because the review process is still within a cycle and the intent was to provide this description for the September CPT meeting.

Recommendation 2. **It is recommended that the model framework description be clearly labelled as such, and placed in a different section to the description of the actual base model used.**

Recommendation 3. **That a table be produced for each test that fully describes the settings and input values used in that test. This should be a direct translation from the main *.dat files that are used by TCSAM02 as input files. This provides clarity on which settings, parameters and data inputs are used for the base case and other updated models.**

The reviewers were also provided with an additional set of detailed runs which did not all conform to tests in the documentation provided. There was, therefore, confusion as to which model the review team would call the current model as described in ToR 1. After some discussion at the workshop, it was agreed that it would be the ~\Google Drive\201707TannerCrabCIEReview\AssessmentModelRuns\BaseModel, also similar to AGO and B0.

An aspect that further increases the level of confusion to an outsider reading the work for the first time is the naming conventions used. The same model and settings can have different names. The same model can have its name changed depending on the report. From the documentation provided, the CPT is busy developing a naming convention system and this undertaking is supported wholeheartedly by this reviewer.

Recommendation 4. Continue to undertake a process of producing a consistent model naming convention that is intuitive and will allow one to recognise the model framework and test. The name for the same model settings should not change over time.

Finally, there are settings abbreviations within the model that are contained in header files and not in the data input files and sometimes not explained in the documentation. These are only known to the assessor (as a whole). However, there would be a benefit in a document that describes how to set up the model and these settings can then be included. This would mean that anyone could truly run this model, where presently there is a great reliance on one assessor.

Recommendation 5. Fully describe how to set up the model code, including tabling the meaning of each of the setting options.

7.1.3 Model mathematics of the current model

7.1.3.1 Maturity

Three forms of maturity are possible in the model: immature, mature new shell, and mature old shell. Mature crabs are assumed to have reached their terminal molt. The mathematics is appropriate.

7.1.3.2 Shell condition

Crabs are modelled whether they are in old or new shell condition, but this factor is not included in any of the likelihood functions. Old shell condition is assumed for shell condition 3 upwards. Old shell crabs are all assumed to be mature and therefore reached their terminal molt, whereas new shell crabs could be either mature or immature.

Given that shell condition is not always an accurate reflection of crab age and does not influence the likelihood, there could be an argument to exclude shell condition entirely. This should not affect the model results. After discussion of this during the review sessions, the case was made that the industry prefers seeing the results in this form, and therefore it seems appropriate to keep shell condition in the model. However, if this does change, some computing time and model complexity could be reduced by excluding shell condition.

Recommendation 6. Consider removing shell condition from the model if there is industry support, given this would simplify the model mathematics and would not directly influence the model results.

7.1.3.3 Recruitment

Although TCSAM02 can assume a sex ratio different to 1:1, the current model uses the 1:1 as a default. A set size frequency for both sexes is assumed. It is unclear where the size frequency data come from.

Recommendation 7. Explain the source of the recruitment size frequency distribution in the documentation.

Recruitment size frequency can be estimated rather than input. Several different approaches are globally used, although drawing from a distribution is more common (whether pre-specified or not). Recruitment to the model is assumed for crabs greater than 25 mm, whereas animals greater than about 125 mm CW enter the fishery and the OFL index are mature males. It is therefore not recommended to change the present system, although direct estimation of the size frequency option and/or using a distribution could be considered at some future date.

Recruitment is estimated in the model using two time blocks with the changeover between 1948 and 1973. The early period is used mostly to set up the model dynamics and is not used in the projections.

For each time block, a mean recruitment parameter and annual deviations around this mean recruitment are estimated. The first period assumes a random walk, whereas the second time block is drawn from a normal distribution. No temporal auto-correlation in the second period is included beyond drawing from a similar distribution. This option should be investigated given discussions (see later in the report) that there may be impacts of environmental drivers on recruitment that might hint at some temporal auto-correlation.

Recommendation 8. Investigate the option of including temporal correlation in recruitment as per Chen et al. (2005); and Punt et al., (2010) (although the latter assumes auto-correlation around a stock recruitment relationship).

Very loose priors are placed on the recruitment deviations, especially for the second time block. Mathematically the setup of the recruitment model, especially the second time block, is reasonably standard for models which do not assume a stock-recruitment relationship.

Recruitment time blocks

- 1948-1973
  - pLnR(1)
  - pDevsLnR(1) ~ AR1(0,1/sqrt(2.0))
- 1974-2015
  - pLnR(2)
  - pDevsLnR(2) ~ N(0,22.4)~N(0,1/sqrt(0.002))

7.1.3.4 Molting, maturity and growth

Tanner crab undertake multiple molts in the first two to three years, but change to being annual thereafter until their terminal molt at maturity. Thus, no growth is assumed after maturity. There was some discussion as to whether skip molting may occur (assumed not possible in the model). Tamone et al. (2007) found that there are hormonal changes at terminal molt providing evidence that the terminal molt does occur in tanner crab and indicates lower likelihood that skip molting occurs.

There is no separate probability of molt to maturity time blocks nor are these set with priors, which is appropriate:

Pr(molt to maturity) time blocks

- males, 1948-2015
  - pvLgtPrM2M(1)
- females, 1948-2015
  - pvLgtPrM2M(2)

Growth for the current model is derived from Kodiak data which are based on data collected from 1994 to 2012. The growth data themselves are not added to the current model, but rather added as priors (see below) in a single time block.

Growth time blocks

- males, 1948-2015
  - pLnGrA(1) ~ N(0.437941, 0.025)
  - pLnGrB(1)~N(0.9487,0.100)
- females, 1948-2015
  - pLnGrA(2) ~ N(0.56560241,0.100)
  - pLnGrB(2)~N(0.9132661,0.025)

TCSAM02 uses the GMACS growth model which assumes the distribution about the mean size after molt in the growth transition matrix follows a gamma distribution. This is a common assumption in many length-based models (although other distributions are also used), unless they have large tagging datasets in which the distribution is estimated directly from the data. The TCSAM02 approach is superior to the TCSAM2013 approximation and should be the default, as is the case for the 2017 assessment and is also supported by the CPT.

Recommendation 9. Agree with the assessor and CPT to use the GMACS growth model as the default approach.

Both models have two sex specific parameters *a* and *b* that are part of the function that models mean size after molt, given the pre-molt size. There is also a scale factor which is not estimated and fixed at 0.75.

In the current model, the female *a* (pLnGrA[2]) parameter hits a bound, despite this model including informative prior weights and the scalar as an input value.

Table 1: Growth parameter values estimated by the current model but with the EBS growth data added (based on the BaseModel+GrowthData directory output files) that come close to or hit their bounds. Min and Max are the minimum and maximum bound settings, value is the estimated value, name references the parameter.

min	max	value	name
0.3	0.6	0.334564	pLnGrA[1]
0.4	0.7	0.663218	pLnGrA[2]
0.7	1.2	0.979823	pLnGrB[1]
0.6	1.2	0.885505	pLnGrB[2]

Tagging data are not included in the current model, so the growth parameters have to be inferred from the size frequency data. This potentially creates conflict with other parameters such as selectivity. Estimating growth without the tagging data, given the complexity of the model, is not ideal and is best avoided. However, this may be a legacy issue in that EBS data have only recently been obtained, whereas the growth priors were based on Kodiak data, the inclusion of which in the model may have created a degree of bias.

Several tests on growth were provided to the previous CPT meeting – comparing the TCSAM2013 and GMACS growth, and including tagging data. Two data sources are available for this – those from the Kodiak data on which the priors are based and the smaller (125 crab) EBS data collected in 2015 and 2016. Results of adding the EBS data and the Kodiak data to TCSAM02 AGO were provided to the May 2017 CPT meeting, which supported the use of the model that includes the EBS data, estimating the scale parameters and using the cumulative gamma function. By including the tagging data, the bounds and priors can be relaxed, which is ideal (but see general comments about priors, bounds, etc. below).

The inclusion of the EBS data removes the issue that the growth parameter bounds are hit, but not the overall issue as several key selectivity and catchability parameters still do so.

Table 2:Growth parameter values estimated by the "BaseModel+GrowthData" model in the directory provided - with the EBS growth data added.

min	max	value	name
0.3	0.6	0.334564	pLnGrA[1]
0.4	0.7	0.663218	pLnGrA[2]
0.7	1.2	0.979823	pLnGrB[1]
0.6	1.2	0.885505	pLnGrB[2]

In terms of growth, the move to estimating the scale parameter is also supported.

In summary, the GMACS growth model with at least the EBS tagging data is essential. The system of provided information growth settings without the tagging data is not supported

Recommendation 10. Agree with the CPT to include the EBS data and free the scale parameter.

7.1.3.5 Commercial selectivity, catchability and retention

Several gear types catch tanner crab, either as a bycatch or by the directed tanner crab fishery. The bycatch tonnage and incidental mortality of the bycatch fisheries can be high and it is therefore essential that these components are well described. However, the complexity of these different gear types is reflected in the wide range of selectivity options available in the model and the flexibility of sharing parameters between functions. Although some vessels may fish in multiple crab fisheries, the quota management systems are distinct and they need different gear to target tanner crab.

A strong aspect of the model is that selectivity parameters are defined independent of the functions themselves, which means that certain parameters can be shared between fisheries (not used in the current model), between mean parameter sets (e.g. used for the SCF) or time blocks (e.g. used for first of the six – pS1- TCF male mean parameter sets across all time blocks). However, this flexibility relies very heavily on the model description being very precise, which is not the case in the documentation provided.

There are several capture/selectivity functions used: ascending logistic (TCF retention, TCF total selectivity females, SCF females, GTF males and females, RKF males and females); ascending logistic with ln(Z50) (TCF total selectivity) and ascending double logistic (SCF males). Upper and lower bounds are applied with no priors, except for the ln-scale Z50 deviations for TCF male selectivity, together with using a random walk prior. All selectivity (except TCF retention selectivities for males) are independent of sex, time block and fishery. Justification for the double logistic of the SCF and not the other crab fisheries is not obviously provided.

The TCF retention selectivity is based on sex-specific ascending logistic functions (females are not retained). Male retention selectivity is parameterized in two time blocks pre- and post-1990 (post when discard data are recorded in detail). Two main parameters define the slope ($p\beta_x^{TCF}$ - beta) and the size at 50% (z50) selection. Total selectivity is also modeled as an ascending logistic function, but with z50 post-1990 calculated through estimated ln-scale mean male size at 50% selectivity and ln-scale deviations in male size selectivity. The z50's are in two time blocks pre- and post-1990/1. The pre-1991 selectivity parameter z50 is related to the z50 parameters estimated for the period 1991-1996, i.e. just after discard data become available. Appropriately, the slope parameter is shared for the period pre-1991 and the period 1991-1996. A separate slope parameter is estimated post-1996. Post-1990, there are ln-scale z50 parameters and annual z50 deviations.

The time blocks relate to when discard data (1991) become available and post rationalization (2005). The assumption that pre-discard data assume similar selectivities to the most recent period post-1990 seems appropriate.

However, it is interesting how much the selectivity function does change from year to year (and model sensitivity test in some cases). On the other hand, growth is very narrowly defined. The TCF catchability values also show large changes just prior to 1980 and around 1996. These outputs show that there are interactions between these parameters and perhaps the transition between blocks. This should be much more thoroughly tested through a comprehensive series of tests. The first would be to free the growth parameters and include the EBS tagging data. The second would be to variably include annual smoothing penalties to see how these changes flow through the model. Despite the needed increase in flexibility in the model, there is a case for considering, as a test, whether penalizing the inter-annual variability in the selectivity deviations and also catchability result in smoother transitions.

Recommendation 11. As a sensitivity test, add a penalty to smooth inter-annual changes in selectivity z50's and catchability, and compare with the current model that includes tagging data.

The same could be considered for all the fishery selectivities.

It is unclear why the female selectivities for the SCF are not also dome shaped. This should be explained. Also the reasons why the RKF and SCF time blocks (which are the same) are different to the TCF.

Recommendation 12. Explain reasons for SCF female selectivity not being dome shaped, given that for the males, the model applies a double logistic function.

Recommendation 13. Explain reasons for SCF and RKF time blocks and why these differ for TCF.

The GTF size frequency data appear to contain more cohort information and samples from a wider distribution of animals. Examine a case where the GTF size-frequency data are more emphasized.

Recommendation 14. Examine a case where the GTF size-frequency data are more emphasized, when compared to the other bycatch fisheries.

All the ascending logistic functions forced the selectivity to be 1 at the maximum size bin of 182. However, the fully selected size is much lower in most of the fisheries, given the data and the parameter estimates. Although this allows for great flexibility in the selectivities, this flexibility appears incongruous. Test the model with lower fully selected values per fishery, based on the data.

Recommendation 15. Test the model with much lower fully selected values per fishery, based on the data.

7.1.3.6 Survey selectivity and catchability

The NMFS EBS Shelf Crab and Groundfish Trawl Survey have been conducted annually from 1975 onwards and use a fixed grid design. There have been some changes to the survey design in terms of its spatial extent, where grids were added during the 1975 to 1982 period and in 1987. The final survey extent is expected to cover the stock well, although females bury, and so are less likely to get caught. Nephrops gear results seem to support this assumption. The survey gear changed in 1982.

Survey time blocks

- Males
  - 1975-1981
    - catchability: plnQ(1)
    - selectivity(1): pS1(1) = z50, pS2(1) = z95-z50
  - 1982+
    - catchability: plnQ(3) ~ N(0.88,0.05)
    - selectivity(3): pS1(3) = z50, pS2(3) = z95-z50
- Females
  - 1975-1981
    - catchability: plnQ(2)
    - selectivity(2): pS1(2) = z50, pS2(2) = z95-z50
  - 1982+
    - catchability: plnQ(4) ~ N(0.88,0.05)
    - selectivity(4): pS1(4) = z50, pS2(4) = z95-z50

The model assumes two time-blocks – pre-1982 and post. The post-1982 period is when the gear changed, more accurate swept area calculations were available, and the spatial extent of the survey was reasonable for tanner crab. This time block therefore, is appropriate. This is also the time block on which projections are based.

Although more stations were added after 1987, they are not likely to have affected tanner crab results, since the post-1979 survey covered much of their distribution. However, the extent of the 1975-1977 survey was much reduced.

The spatial coverage changes are not reflected in the survey catchability (q) or selectivity time blocks (below), whereas the gear change is. Most of the spatial change is before 1977/8, which would be difficult to separate as a time block (except starting the model after 1978, which is unnecessary). The input sample sizes for the NMFS trawl survey is 200. Male selectivity is assumed to follow an ascending logistic function (using the alternative parametrization of z50 and Δz95-z50 i.e. the latter is the difference between the size at 95% and 50% selectivity) with two time blocks, pre-1982 and 1982+. Given the change in early survey design, investigate the option of adding another time block for the survey parameters. As can be seen (and discussed further below) in Table 3, several of the survey selectivity and catchability parameters hit their bounds – z50-z95 (females), lower bounds for males and female catchabilities for the 1975-1981 time block. This means the catchability and selectivity functions for the early time block are not well specified. The survey catchability parameters remain on the lower bound even in the test where the model bounds and priors are weakened. The flexibility of the selectivities for the fisheries are already unclear, but here there is no reason why the upper selectivity would change from year to year. Furthermore, the size at full selection can now be obtained using the side-by-side surveys. For this reason, an alternative selectivity function fixing the fully selected (99% selectivity) is recommended.

Recommendation 16. Fix the size of 99% selectivity based on size-frequency data and/or the side-by-side trawl data and only estimate z50.

This seems similar to a suggestion post-review by the assessor in an email. An alternative to the above is to assume a gamma distribution. This has often been shown to fit the data better than the logistic functions (e.g. Deng et al., 2015).

7.1.3.7 Natural mortality

In the current model, a base level of natural mortality is input with estimated deviations from the base level for (immature and mature) male and females respectively. The base natural mortality value is calculated from the Hoenig approach with the assumption that crab longevity is 20 years. This longevity of 20 years is based on snow crab data. This calculation fell within the range of catch curve analyses undertaken by Somerton (1981). The base priors in the main period are offset with very informative priors. Given that there is more uncertainty in the base M than implied by the prior, a less informative prior should be tested.

Recommendation 17. Test a model with less informative priors for the base M.

Natural mortality time blocks

- Immature crab: 1948-2015
  - pDM1(1) ~ N(1,0.05)

- Mature males
  - 1948-1979, 1985-2015
    - pDM1(2) ~ N(1,0.05)
  - 1980-1984
    - pDM1(2) ~ N(1,0.05)
    - pDM2(1)

- Mature females
  - 1948-1979, 1985-2015
    - pDM1(3) ~ N(1,0.05)
  - 1980-1984
    - pDM1(3) ~ N(1,0.05)
    - pDM2(2)

An additional time block is added between 1980 and 1984. The enhanced M (EM) time block has no prior, i.e. it can be the same or different to the base prior in any direction. This uninformative EM prior system is an appropriate assumption.

The need for the EM time block was first highlighted because of consistent lack of fit to the data over this period. Studies showed mortality from cod predation over that period was very high, since cod abundance was very high. The disconnect between the model and the data was assumed to be due to additional natural mortality. In most tests, the EM is much increased compared to the base M. However, this result does not totally support the EM hypothesis, as other factors such as survey catchability are not freed over the same period. An interesting test to the EM hypothesis would be to add a survey q block for the same period. Even though there is no difference to the survey design, the side-by-side trawl experiments have indicated that survey q can change.

Recommendation 18. Test whether a survey q block over the same period changes the EM results.

Although there is some doubt on the EM view given the recent side-by-side trawl results, the hypothesis of increased mortality is more likely to be true. However, there is some inconsistency in the time period used, as cod biomass values provided during the review seem to remain high, arguably from 1980 to

1987 (Figure 1). Extending the time block to better reflect cod biomass should be tested. It should also be noted that recent biomass trends are approaching similar scales.

Recommendation 19. Test extending the EM period to 1980-1987.

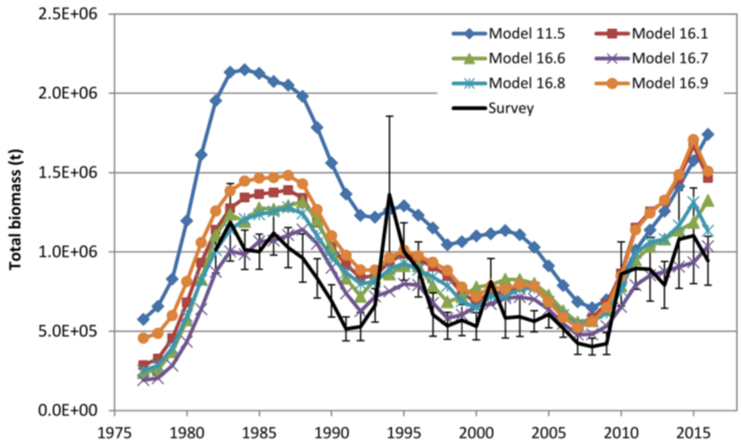

Figure 1: Pacific cod biomass estimates provided during the review in 2. TannerCrabBiology.pptx.

No time block is assumed for immature crab, i.e. the assumed increased mortality is only applied to mature males and females. This seems strange given predation is likely to have been as much or more on smaller, soft crab, which may not have been well represented in the data.

Recommendation 20. Test including a time block for immature crab.

However, the *ad hoc* nature of this solution does lead one to consider using a less blunt approach. Although the data may not support this approach, constrained 5-yearly (or smoothed annual) M deviates (with less constraint over the EM) period may be more attractive.

Recommendation 21. Investigate, the use of 5-yearly (or smoothed annual) M deviates.

7.1.4 Retrospective analyses

Ten-year retrospective runs were undertaken for the current (base) model during the review. The MMB shows reasonable stability retrospectively, especially after 1982 (Figure 2). Recruitment is provided in Figure 3. There is some sign in the model of overshooting or undershooting R and MMB over a transition period, but this is the time where little information about recent recruitment is available.

One would assume some of this stability in the retrospective results may also be due to informative priors, bounds, weights, etc. in the model. (See comment in ToR 6 – Other types of retrospective analysis that could be undertaken).

Mature biomass

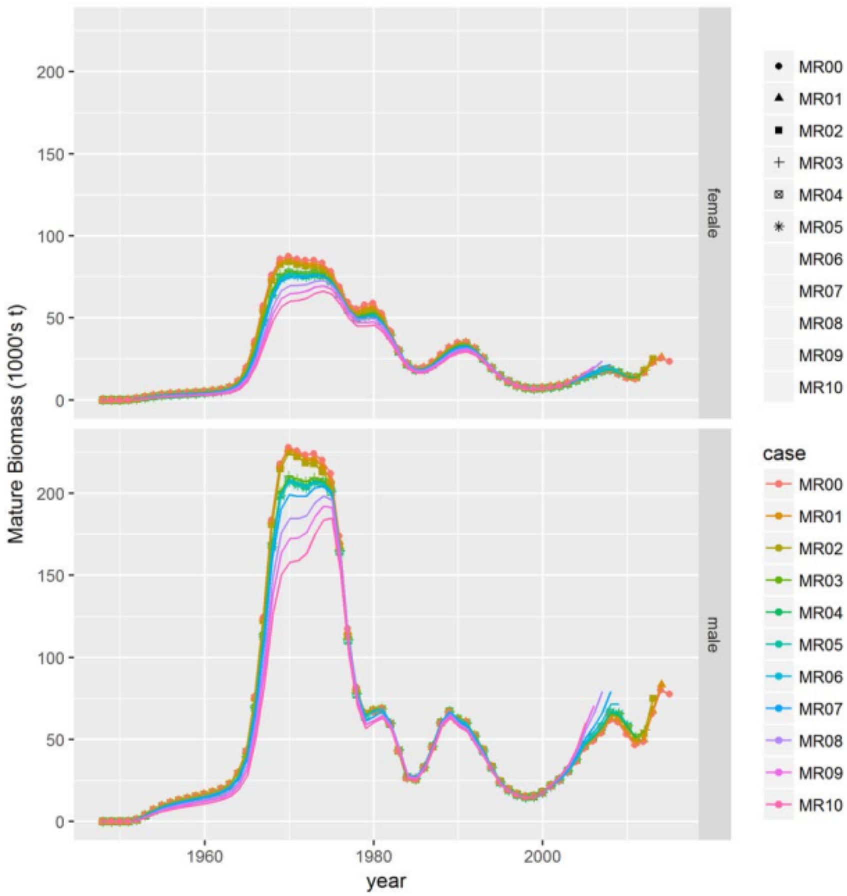

Figure 2: Estimated annual mature biomass from retrospective analyses.

Recruitment

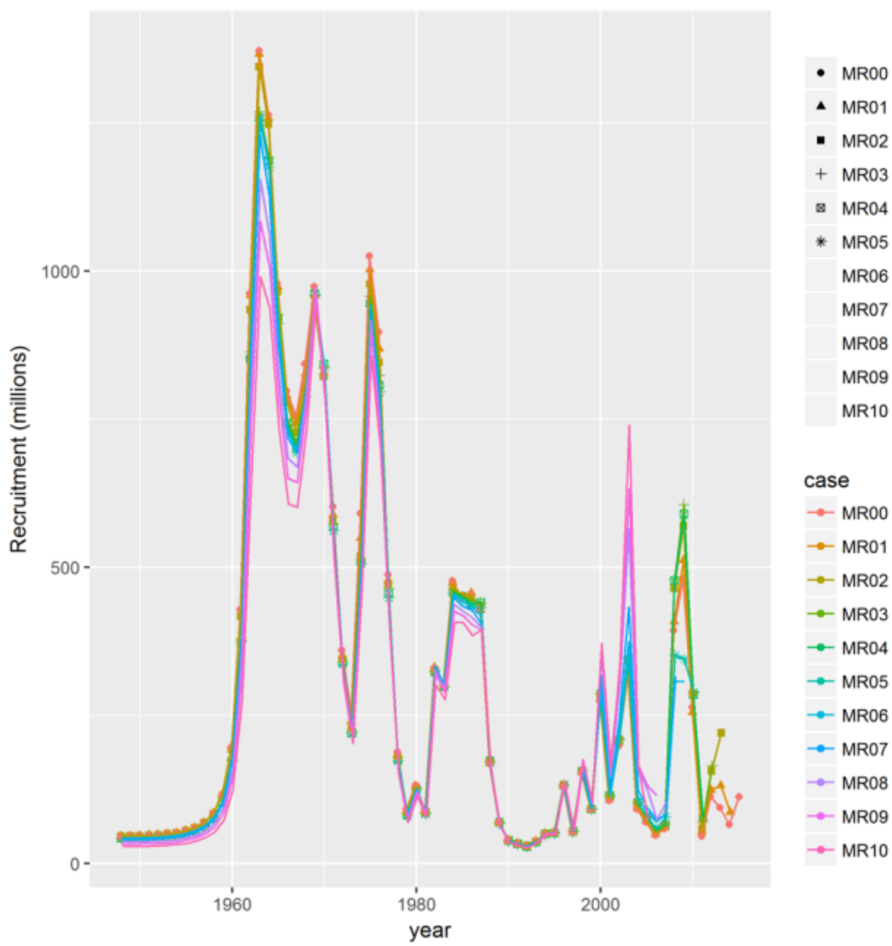

Figure 3: Estimated annual recruitments from retrospective analyses.

7.1.5 Bounds/priors/weights/penalties/sample sizes

A key weakness in the current model is the complexity and degree to which the model is constrained. Several types of systems are applied in the same model to constrain the solution – sample sizes, parameter priors, parameter bounds, likelihood weights and penalties. In many cases, informative priors are used. **This set up is not supported.** It is essential that the model is simplified and modified to be less constrained. Sensitivity tests have shown that some very large changes are possible with small changes in the priors/ penalties etc., model structure or data inclusion. The fundamental reason for this needs to be understood. It could be model misspecification and/or conflicting data. See ToR 6 and 7 on steps to rectify this.

The current model is highly constrained. Key issues are:

1. Several key parameters hit or are close to their several bounds, for the current model, these are shown in Table 3. They include the *a* female growth coefficient (pLnGrA[2]); the male and female logit-scale parameter vectors for Pr(maturity-at-size); some of the selectivity parameters (in particular the

slope *z50* parameter) in different year blocks for the groundfishery (GF), the red king crab fishery (RKF), the snow crab fishery (SCF) and the NMFS survey; and the male and female survey catchability parameters in the first time block. All these parameters, except the female growth coefficient, are bounded without priors.

pLnGrA[2] is bounded between 0.4 and 0.7, with the initial value set essentially at the upper bound(based on the best jitter run) and a prior from the normal distribution with a mean and variance of 0.56560241 and 0.100 respectively; and a prior weight of 1. As a result, this parameter is started at the bounds and remains at this value despite the prior mean being well below the initial and final value. <u>This highlights a conflict between the growths and other parameters.</u>

Table 3: Parameter values estimated by the current model that come close to or hit their bounds. Min and Max are the minimum and maximum settings, value is the estimated value, the process and label references the parameter.

min	max	value	Code name	type	process	label
0.4	0.7	0.6999999	pLnGrA[2]	'param_init_bounded_number'	'growth'	'females'
-15	15	14.99999911	pLgtPrM2M[1]	'param_init_bounded_vector'	'maturity'	'males_(entire_model_period)'
-15	15	-14.99999969	pLgtPrM2M[2]	'param_init_bounded_vector'	'maturity'	'females_(entire_model_period)'
40	250	40.00000001	pS1[20]	'param_init_bounded_number'	'selectivity'	'z50_for_GF.AllGear_selectivity_(males_1987-1996)'
95	150	149.9999992	pS1[23]	'param_init_bounded_number'	'selectivity'	'z50_for_RKF_selectivity_(males_1997-2004)'
95	150	150	pS1[24]	'param_init_bounded_number'	'selectivity'	'z50_for_RKF_selectivity_(males_2005+)'
0	100	99.99999986	pS2[4]	'param_init_bounded_number'	'selectivity'	'z95-z50_for_NMFS_survey_selectivity_(females
0.1	0.5	0.499999038	pS4[1]	'param_init_bounded_number'	'selectivity'	'descending_slope_for_SCF_selectivity_(males_pre-1997)'
-0.69315	0.001	-0.693147149	pLnQ[1]	'param_init_bounded_number'	'surveys'	'NMFS_trawl_survey:_males_1975-1981'
-0.69315	0.001	-0.693147051	pLnQ[3]	'param_init_bounded_number'	'surveys'	'NMFS_trawl_survey:_females_1975-1981'

2. The likelihood profile for derived parameters *may* be narrow and highly specified

Likelihood profiles of MMB and other derived parameters were requested and provided. These showed very narrow posteriors given the priors. **However, since the review workshop it has been discovered that the recent versions of ADMB have a bug and these posteriors are likely to be incorrect. Other completely different models tested since the review are showing similar results. These likelihood profile results obtained during the workshop are therefore not presented.**

Recommendation 22. It is recommended that once this bug has been removed from ADMB, that the current model likelihood profile gets recalculated and becomes part of the report to the next CPT (High priority).

3. The jittering results show that the model does not converge to the same minimum value and includes some optimizations that are quite divergent from most of the other runs (Figure 4). One would assume that the few extreme jitter runs are due to combinations of starting values that do not make biological sense and clearly should be discarded (as done). However, the reason for these outliers should be checked.

Given the results below, questions arise as to whether the solution of the jittering even found the global minimum, i.e. would 200 more jitter runs sampled from a wider range of possible starting values have produced a different “best” model. This problem is not as pronounced for the current model, compared to one of the scenarios shown at the workshop, where the jitter results created the impression that there were separate biological states described by the results, depending on the starting values.

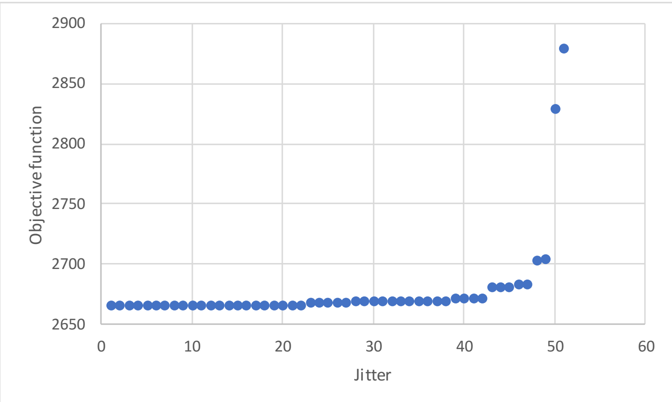

Figure 4: Objective function sorted in ascending order for the above subset of the current model jitter results.

The current model jitter results in terms of final year Mature Male Biomass (MMB) (Figure 5) shows that the range of possible final year MMB values appears to lie within the range of 73 and 78 million pounds for similar objective function values.

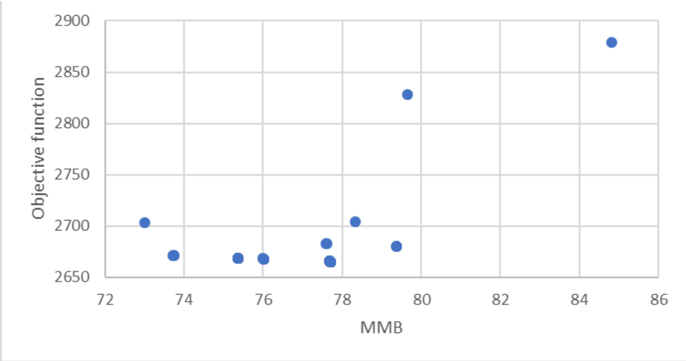

Figure 5: Objective function from a subset of the jitter results related to the MMB for the current model. Note there are several points that overlap.

Recommendation 23. It is essential that the causes of this behavior are investigated, rather than restricting the model further or only selecting the best model.

4. A sensitivity test was undertaken during the review that increased the bounds and some priors for the current model. The differences between the base and this test show that this model is sensitive to these input settings, e.g. M parameters (where the EM values are less than the current model's and the base values higher - Figure 6), total biomass, etc.

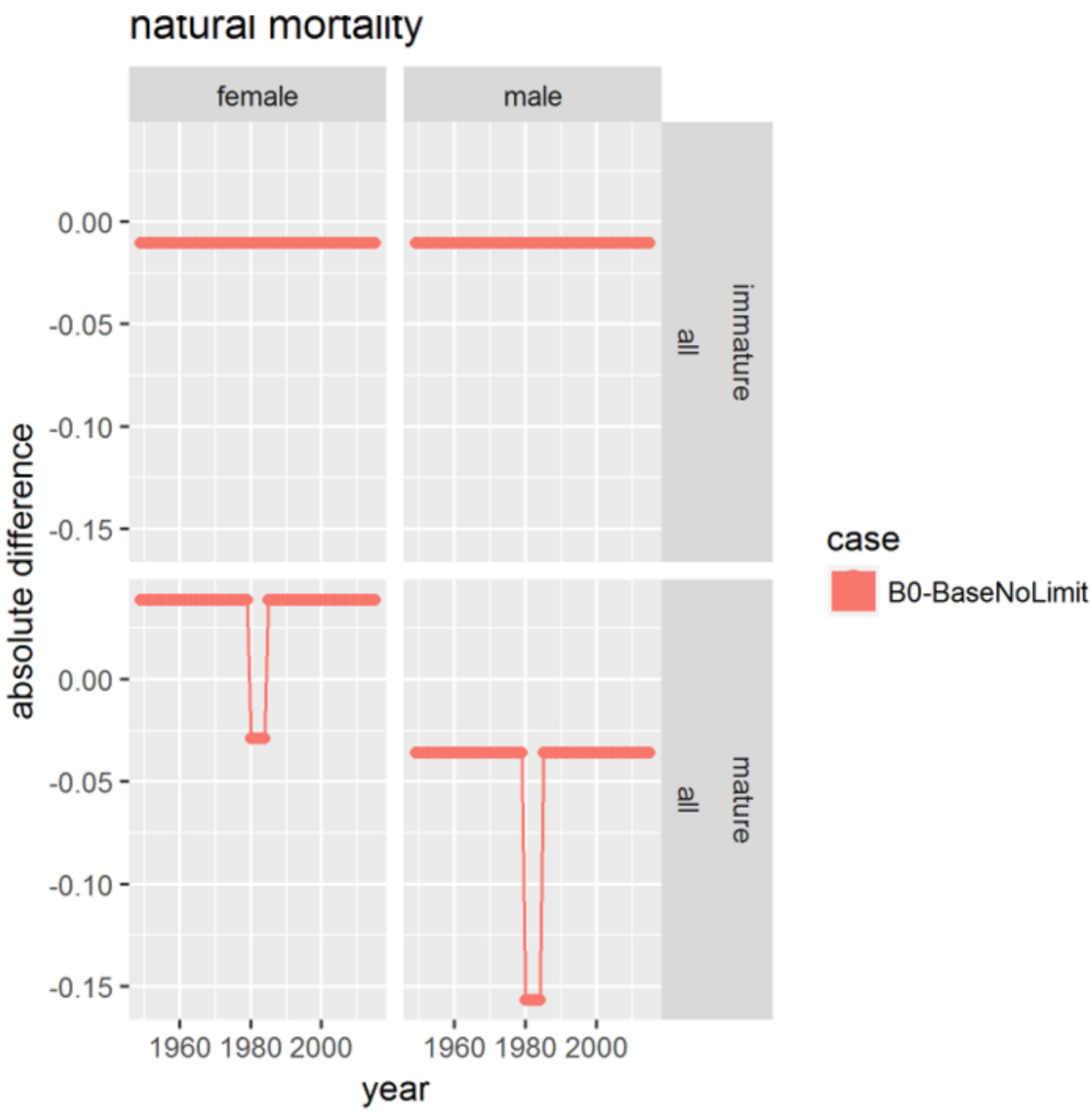

Figure 6: Difference in estimated M parameter values between the current B0 model and the No Limits model.

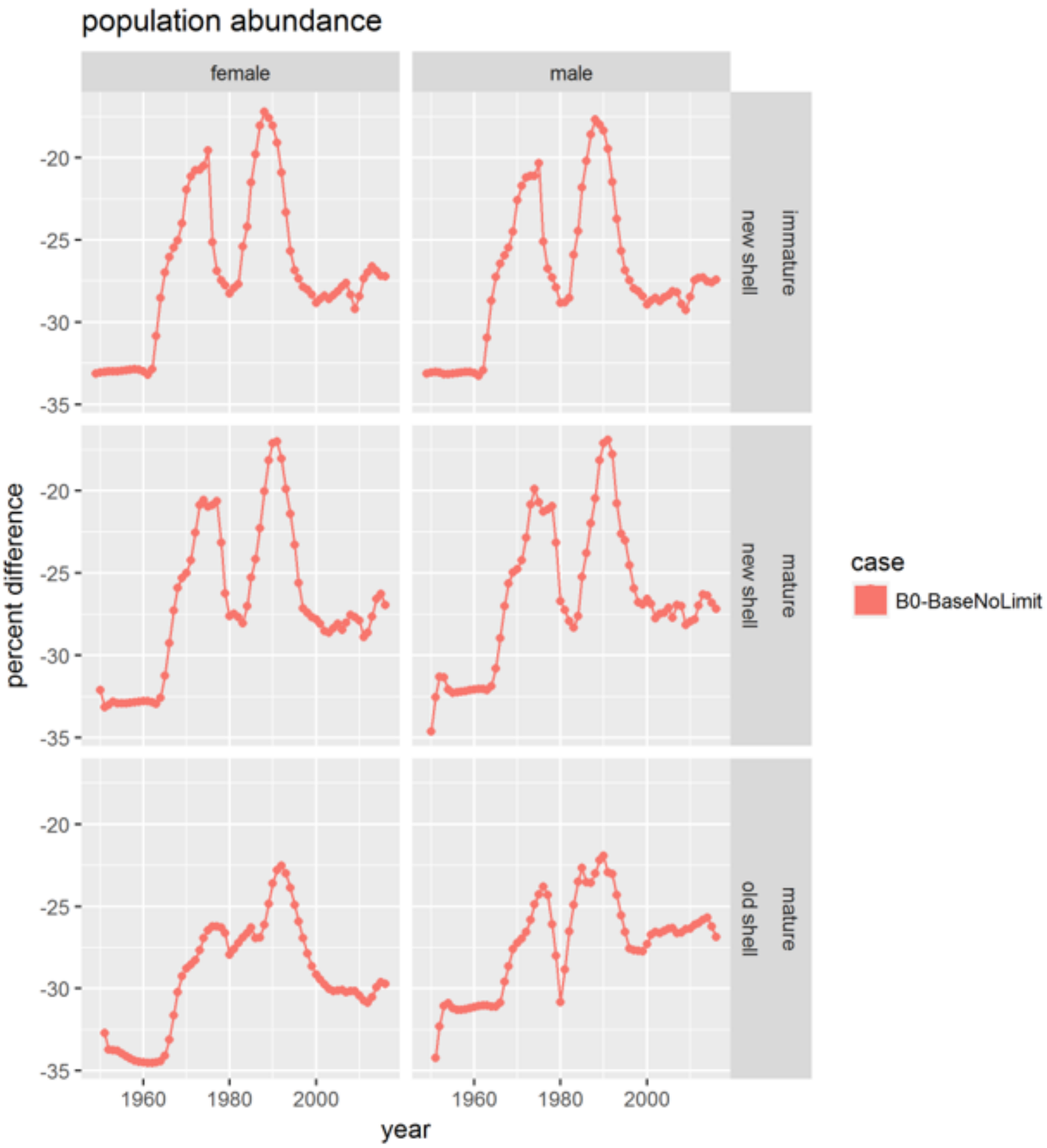

Figure 7: Differences between the current base model (B0 and the No limit test).

Fewer parameters hit their bounds compared to the current model, but it can be observed that new ones also occur. This shows (not unexpectly) that a far more subtle approach to freeing up the model is required (Table 4). This aspect and associated recommendations are discussed further in ToR 6.

Table 4: Parameter values estimated by the current model without limits that come close to or hit their bounds. Min and Max are the minimum and maximum settings, value is the estimated value, the process and label references the parameter. Parameters in italics also hit their bounds for the current model.

min	**max**	**value**	**name**	**type**
-50	69	-49.994826	pS1[4]	z50_for_NMFS_survey_selectivity_(females,_1982+)
95	*250*	*249.999909*	*pS1[23]*	*z50_for_RKF_selectivity_(males,_1997-2004)*
95	*250*	*249.999994*	*pS1[24]*	*z50_for_RKF_selectivity_(males,_2005+)*
0.25	2	0.2531387	pS2[6]	slope_for_TCF_retention_(1997+)
-2	*0.0009995*	*-1.9518057*	*pLnQ[1]*	*NMFS_trawl_survey:_males,_1975-1981*
-2	*0.0009995*	*-1.9999999*	*pLnQ[3]*	*NMFS_trawl_survey:_females,_1975-1981*
-2	0	-2	pLnQ[4]	NMFS_trawl_survey:_females,_1982+

The sensitivity run has highlighted some important aspects:

1. The objective function for the model without limits is lower than that for the base model, although some of the parameters do not appear to be realistic.

2. That a large range of model results is possible given the restrictions on the model are influential. Without truly understanding the influence and reasons for the results, a specific model output should always be questioned.

3. It is possible to eliminate some of the constraints being imposed (a simple first extreme run managed to converge). The ideal is a much less constrained model, where removing the constraints is more fully explained and the model converges to the same global minimum during jittering.

7.1.6 Likelihood functions

The likelihood components discussion seemed more appropriate in ToR 3, where it is discussed together with the data.

7.1.7 Summary

1. The TCSAM02 framework is supported. It is flexible and able to undertake additional tests without changing the code. The input files (once understood) are reasonably intuitive.

2. The documentation provided on TCSAM02 is reasonably complete and clear. However, there is some confusion as to what the current model settings are in the documentation. Some of this confusion occurs when the assessor attempts to describe both the framework and current assessment together. Separating these two components should help the reader.

3. The mathematics in the model, in the main, are appropriate, especially the recent inclusion of GMACS components.

4. The model appears to be too constrained and uses too many different approaches for this process. Solutions are not unique and are overly dependent on input settings. A suggested way forward is provided in ToR 6.

7.2 ToR 2: Statements assessing the strengths and weaknesses of the current Tanner crab stock projection model, with regard to methodology.

The assessment projections are presently undertaken within the assessment rather than as separate code, as was the case previously. This approach is a great step forward compared to undertaking the projections in separate code, as the reviewer has seen on occasions where the handshake between the two models was error prone. The projection is used to calculate the OFL, ABC, and stock status based on Male Mature Biomass (MMB). An analytical solution assuming deterministic equilibrium dynamics is used. This has been compared to simply forward projecting the model long term until equilibrium is achieved and comparing the two sets of results. This test demonstrated that the two approaches were similar (not surprisingly, but a good bug test).

However, this is a deterministic projection and should be compared with that of a stochastic forward projection. The norm is that slightly different answers are observed, but the degree of these changes can be variable. With length-based models this can be computer resource intensive.

Recommendation 24. Undertake a stochastic projection to investigate the appropriateness of assuming deterministic projections.

The tanner crab assessment falls into the tier 3 category for OFL calculations, as it has no reliable stock-recruitment curve. As a result, $F_{35\%}$ and $B_{35\%}$ are used for F_{MSY} and B_{MSY} proxies, respectively. In the tanner crab fishery, the biomass is based on MMB as it is a male-only fishery.

The projection is performed for one year using average recruitment over the period from 1982. Importantly, and correctly, the bycatch mortality is included in the OFL calculation. In the case of the bycatch fisheries (except the SCF), the most recent 5 years' settings are used in the projection. In this case, the actual OFL set for the snow crab fishery in the assessment projection year is used. These settings are appropriate given that the fishing mortality impact on the SCF is usually larger than the other bycatch fisheries.

There is no stock-recruitment relationship assumed or shown to occur in the model. Breakpoint analyses undertaken in the past proposed that there may have been a regime shift in the stock-recruitment results after 1985. However, the evidence for this was deemed weak and not supported. In the OFL calculations, the forward projections are based on recruitment from 1982 onwards. This is the period after which there was stability in the NMFS groundfish survey in terms of gear and spatial extent. However, this does include the period where there is an *ad hoc* increase in the assumed natural mortality (enhanced mortality). For this reason, it may be useful to include a sensitivity test where more recent recruitment years are used in the projection (for example, from 1990 after the high cod abundance years).

Recommendation 25. Undertake a sensitivity test using more recent recruitment years in the projection.

The distribution of the OFL can also be obtained using MCMC. This may be less useful with such a constrained model, but may be more so with later versions.

Recommendation 26. Undertake MCMC to calculate the distribution around the OFL.

Similarly, the projection uses the average recruitment since 1982, which again does not fully test the uncertainty in the OFL calculations. The distribution in the recruitments in this period can be sampled in the forward projections.

7.2.1 Summary

In summary, the projections are undertaken appropriately. Stochastic versions and choosing a more recent time period for sensitivity tests should be undertaken.

7.3 TOR 3: A REVIEW OF THE FISHERY DEPENDENT AND INDEPENDENT DATA INPUTS TO THE STOCK ASSESSMENT WITH REGARD TO QUALITY OF INFORMATION AND APPROPRIATENESS TO THE ASSESSMENT.

7.3.1 NMFS Surveys Abundance Index

The NMFS EBS Shelf Crab and Groundfish Trawl Surveys have been conducted annually from 1975 onwards, using a fixed grid design. There have been several changes over the early period, but the design and method used has been substantially stable after 1987. The model uses aggregate catch biomass divided by sex only. The data are aggregated using the standard statistical approach; scaling over size, site, and time, and this is appropriate. The aggregation has been undertaken by both the assessors and independently through the survey data report and are in agreement. The annual means and associated actual (not effective) variances are provided to the model and the likelihood assumes a lognormal function with a weight of 1.

These surveys have also contributed to the basic understanding of crabs, amongst others. Without this dataset, the model would need to rely on other abundance indices such as catch rate, which would raise substantial issues of standardization.

The use and application of these data is appropriate and fundamental.

7.3.2 Commercial landings and effort

In the current model there are four fisheries, being the directed Tanner crab fishery (TCF), snow crab bycatch fishery (SCF), groundfish trawl bycatch fishery (GTF), and Bristol Bay red king crab fishery (RKF).

Retained catch in terms of abundance, biomass and effort comes from fish ticket data. Dockside samplers also record landed size composition for vessels without onboard observers and obtain landing mass. In addition, ADFG crab observers sample the size composition on the floating processors and at-sea observers on crabbers (directed, snow crab and RKF fisheries). The trawl fleet has 200% observer coverage, crab observer coverage is 100% on processors and about 30% of fishers. Observers are randomly assigned a sub-set of the vessels and these are sampled through the season. Based on discussions in the workshop, the 30% observer coverage provides true full season coverage, whereas in some fisheries these can be 30% of the vessels for a single of many trips only, i.e. the actual observer coverage is much lower. This level should appropriately sample the fishery, particularly in terms of biomass and size frequency.

Retained catch data can be obtained from the fish ticket data and the combination of dockside and at-sea observers.

For estimates of total catch abundance or biomass, the assessor uses the average weight inferred from the size frequency and length-weight regressions. However, the data are pooled without adjustment for the difference in the size composition for the east and west regions. Average weight is calculated from the size frequency. Although this is usually a good approximation, it would be worth, as part of a simulation study, to investigate any biases obtained from this approach, whether through regional differences or the use of average weight, rather than using the full weight frequency from the size distribution.

7.3.2.1 Foreign fleet

There are also two historical foreign commercial fisheries: Japanese and Russian, not treated as separate fleets. There is some uncertainty surrounding these data, since investigation of the reports had to split "tanner crab" into tanner and snow crab. Historically, most of the recorded tanner crab would have been tanner crab. For this reason, the uncertainty is taken to be small and not considered in the model. Discarding in the early part of the data is more uncertain compared to the catch, which is more certain and a standard deviation of ½ MT is assumed.

However, large under-reporting errors in the foreign fleets' data have been shown elsewhere in the world. A sensitivity test therefore was undertaken that assumed the catch of these two fleets were underreported by 20% with the idea that, apart from the immediate effects on population size, these may explain some of the need to account for the additional mortality in the early 1980s. However, in the results provided during the review it is clear that:

a. there are large (as expected) differences, compared to the current base model, between the population sizes in the recruitment in the early and later parts of the series, but these changes stabilize by the 1970s (Figure 8).

b. Natural mortality estimates remain essentially unaffected (Figure 9).

Any under-reporting of the foreign fleet catches at this tested scale would not affect the major estimates post-1982 of the model, and therefore not the projections. Unless there is some indication that the scale of under- or over-reporting of the foreign catches is substantial, emphasis on correcting these data are not a priority.

Recommendation 27. It may be more appropriate to increase the standard deviations assumed for years where the foreign fleets were active, or enter the data into the model as a separate (uncertain) fleet.

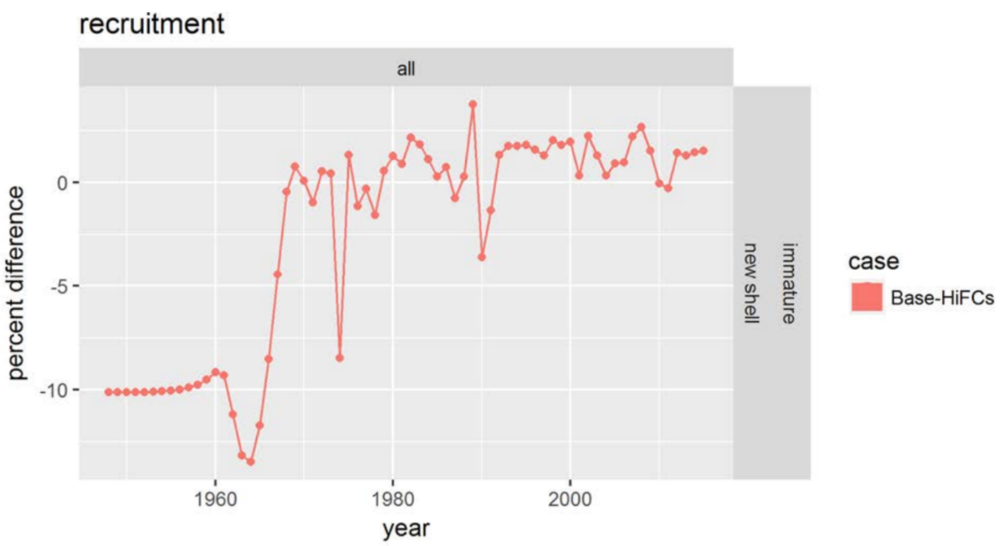

Figure 8: Percent difference in recruitment between the current (base) model and the scenario where 20% more catch was added to the foreign catches.

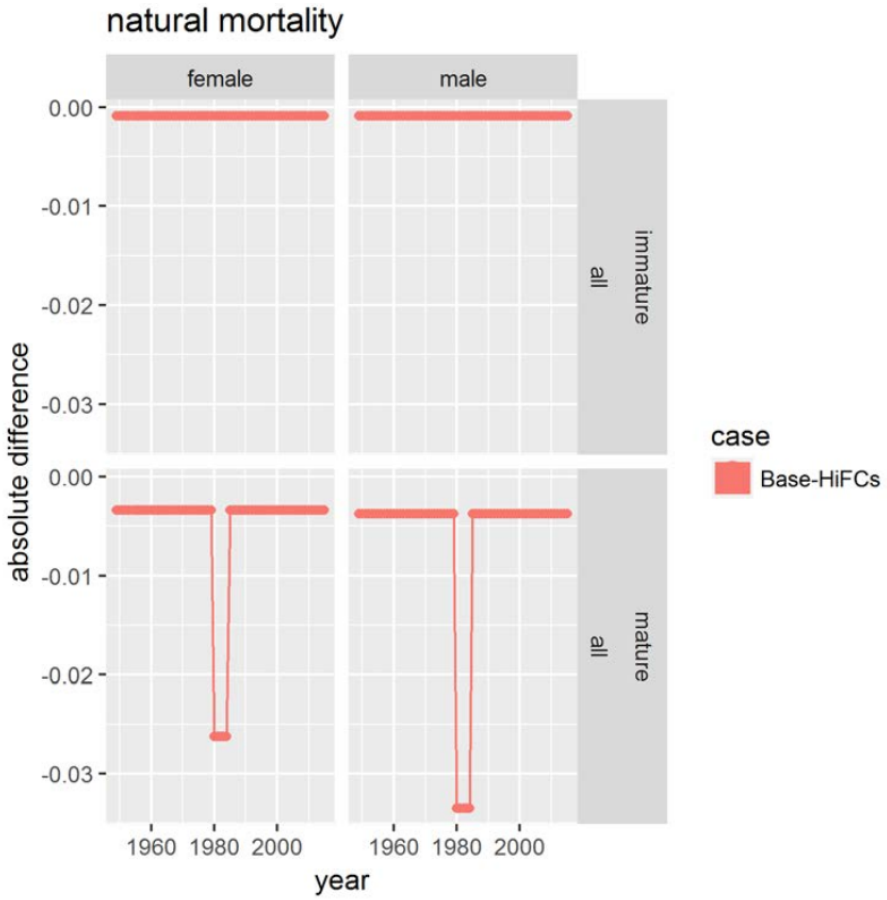

Figure 9: Absolute difference in male and female natural mortalities from the base and the scenario where 20% more catch was added to the foreign catches.

7.3.2.2 Groundfish trawl bycatch fishery

The ground fishery catch is entered in the model as total catch (aggregated for all sexes, shell conditions and sizes) with a constant annual input assumed CV of 0.2 (assumed for all the bycatch fisheries). The likelihood assumes a standard normal likelihood with a variance of 1 and with a likelihood weight of 20. Discards in the ground fishery up to 1991 were combined sex biomass and sex-specific composition. After 1991 this was differentiated by gear. The scaling up of these data during the workshop was described to be consistent with the above approach. Tests were undertaken to separate these.

No effort data are used in this fishery.

7.3.2.3 Crab fisheries

The crab bycatch fisheries' catches are entered as aggregate catch with the same input constant assumed CV and a standard normal likelihood weight of 20. This catch is divided into males and females.

Discards in the TCF fishery are divided into two components: prior to 1991 where discards are not directly sampled, and 1991+ is sampled in terms of sex and area-specific abundance, biomass, size compositions, and shell condition. A similar issue occurs in the SCF and the RKF fisheries, although here tanner crab cannot be retained.

Uncertainty in the discard values is higher in the earlier part of the data series. For example, these were not recorded in the earlier part of the data and later discard data are used to extrapolate discards in the earlier period. Unstandardised annual effort data are therefore used to extend the predictions of the discard mortality rate prior to 1992. It assumes that the ratio of the fishing mortality rate per pot lift has remained constant over time. The GMACS approach to modelling fishery capture rate and the associated likelihood function assumes a standard normal likelihood with a weight of 1. In effect, this approach seems reasonable given that no additional data seems available that could provide more information on whether this ratio has changed over time, other than to investigate how variable this ratio over time for the years is and whether there is an annual trend in the ratio.

Recommendation 28. Investigate the present bycatch crab fishery discard data for trends over time.

7.3.3 Size frequencies

7.3.3.1 Fishery specific size frequencies

Commercial GTF and RKF size frequencies are entered into the model, aggregated to male and female sizes. A multinomial likelihood is assumed with a likelihood weight of 1. The input *actual* (not effective) sample sizes for the fishery based size frequency data are calculated using the below approach, which compares the sample size for a sex and year within a fishery to that of the average retained crab sampled by the dockside observers. This is a rather *ad hoc* approach, further discussed in ToR 6/7.

Input sample sizes for size compositions

- NMFS trawl survey: 200
- Fisheries

$n_{f,x}^{y}$ is the number of sex x crab measured in fishery f in year y.

$$\bar{n}_{TCFR} = \frac{1}{N}\sum_{y} n_{TCFR}^{y}$$ average number of retained crab sampled by dockside observers

$$s_{f,x}^{y} = \begin{cases} s_{min} & if\ \frac{n_{f,x}^{y}}{\bar{n}_{TCFR}} \le s_{min} \\ \frac{n_{f,x}^{y}}{\bar{n}_{TCFR}} & if\ s_{min} \le \frac{n_{f,x}^{y}}{\bar{n}_{TCFR}} \le s_{max} \\ s_{max} & if\ s_{max} \le \frac{n_{f,x}^{y}}{\bar{n}_{TCFR}} \end{cases}$$

input sample sizes

$s_{min} = 0$ and $s_{max} = 200$

The size frequency data for the fishery are estimated using a multinomial likelihood with a weight of 1. The weight units are not scaled. Suggestions for alternative approaches are provided in ToR 6/7.

The method of aggregating the size frequency data, scaled up within vessel, size, maturity, etc. for each area, expanded for total area and combined is appropriate. This means that the statistical properties of the sampling are retained.

The observers select a random pot prior to sorting and the species composition, sex and legal status of tanner and other important crabs are collected. For a subset of these sampled pots, further information about size, sex, shell condition, maturity, etc. are collected.

The observer does not keep track of the sampling location, but each sample does have the latitude and longitude, which means this could be inferred if required.

The TCF discard size frequency and volume is assumed from the observer data and not directly recorded post sorting by the crew. This is for practical reasons, but does mean that any shift in the selectivity of the fleet will be reliant on the difference between the observer and fish ticket data. Dockside observers do not resample observed vessels which could have been used to check the retained assumptions of the at sea observers.

Observer training is well established and the dockside sampler has the same training as the at sea observers.

Discards since 1991 in the ground fishery can now be included by gear type. This is because fixed gear bycatch has decreased markedly since the start of the series and is approximately similar to the trawl catch in recent years. Both are now a small bycatch compared to previously. The approach of separating the gear types is supported.

7.3.4 Model framework

TCSAM02 is able to aggregate and scale up abundance, biomasses and size frequency at various levels of aggregation. However, this is probably not the best approach to use, given it is unclear whether there are any questionable results during that process due to outliers, etc. Also, combinations could be selected by the novice user that do not make sense. As a result, it is better that the scaling up and aggregation process, including error calculations are controlled outside the model and these data are then entered in the model. This is the approach currently used.

Recommendation 29. Scale, aggregate and calculate errors outside of the model so that any outliers, data errors and other spurious results can be investigated external to the model, as these would be more evident.

7.3.5 Summary

1. The tanner crab assessment is data rich, although complicated in that tanner crab are caught as a bycatch in several fisheries as well as the directed fishery.

2. Survey data are extensive and cover most of the spatial range of the species, especially after 1987.

3. Total catch, discard and size frequencies are obtained from the fleets either through at-sea observers or dockside.

4. Observer coverage is good.

5. Classic statistical (as opposed to model based) aggregation techniques are used to the highest standard. As a whole, there do not seem to be major issues with regard to the quality of the data and they are generally appropriately used.

6. Despite this statement, some improvements can be made, although most of these are discussed in ToR 6/7 with regard to simulation testing and with regard to constructing a different approach to calculating effective sample sizes and the model estimation process.

7. Potential uncertainty in the foreign catch data was tested and showed not to have large impacts to the post-1982 model outputs. This uncertainty is therefore not a priority.

7.4 ToR 4: Recommendations for alternative approaches to evaluate model convergence and compare multiple models.

7.4.1 Diagnostic tests

The CPT has produced a very rigorous set of guidelines (prepared by Punt and Kinzey) of which diagnostics and tests should be provided. Those provided are supported.

Diagnostic tests include residual plots, marginal observed and predicted plots, observed or input and implied effective sample size. Sensitivity tests include classic sensitivities to likelihood weights for the MMB projections; alternative models changing one assumption or dataset at a time and retrospective analyses. Furthermore, comparison of objective function components has also been provided. These have been comprehensive, but perhaps could be laid out in the documentation more consistently (again, this could be because the review is within cycle).

7.4.2 Retrospective analyses

Retrospective analyses have tended to only fall in one class, which is to test sequentially dropping one year of data at a time. What has not been provided, is how the different models used to set the OFL have changed over time. Given the structural changes that have occurred in this assessment, it is expected that these changes are likely to be far greater than the first form of retrospective analysis.

Recommendation 30. Undertake a retrospective analysis that shows the different assessment models used to set the OFL over time.

Another form of retrospective analysis, is to test how well projections have been undertaken. There are two ways. In both, projections are undertaken retrospectively. The one could be again with the same model moving backwards and the other undertaken of each year's assessment. The predicted data are then compared with the actual. Predict the data – go back a few years and predict the data again (not something you estimate or calculate).

Recommendation 31. Undertake retrospective projection analyses.

7.4.3 Profile likelihoods

Profile likelihoods are not provided as a norm. These are important diagnostic tests for checking the performance of a model, and this should be undertaken not only for parameters that are estimated, but also for derived parameters such as MMB. Unfortunately, the bug in ADMB (which was highlighted directly after this review) meant we cannot show the results of runs undertaken during the workshop, but the approach was demonstrated.

Recommendation 32. Undertake profile likelihoods of key models when the ADMB bug has been removed.

7.4.4 Jittering

Testing for model convergence in terms of global versus local minima has been undertaken through jittering, which is an appropriate process to undertake. In the current model, 200 jitter runs are undertaken without running the full model (e.g. without projections and OFL calculations). The run with the lowest objective function is chosen as the best model. This model is then run fully (including the projections) based on the selected lowest run's initial values. The jitter results are not closely investigated beyond finding the lowest value, nor were important values such as MMB stored prior to the workshop. The range of the jitter values are an input and the default is from 20% to 80% of the parameters' ranges.

During the workshop, the current model jitter results were re-run and important values such as MMB were stored. These results are discussed in ToR 1, but should be investigated more closely as standard procedure. Ideally a model should converge to the same solution no matter what the initial values are. The fact that there are local minima (in some cases) or a somewhat flat likelihood surface highlights that this phenomenon should be more closely investigated. Also, a greater number of jitter runs should be undertaken in these circumstances as well as sampling from a larger range of values, despite increased computing time requirements.

Recommendation 33. Investigate the jitter results more fully as standard procedure and retain all results.

Recommendation 34. Undertake a greater number of jitter runs, sampling from a larger range of values.

Recommendation 35. Investigate the probable causes of these local minima more fully (see simulation and MCMC sections below, and ToR 6). – HIGH PRIORITY

AIC still remains a good approach to make statistical comparisons between models, but should be used in combination with all of the above.

7.4.5 MCMC

MCMC is a very useful approach for estimating the parameters and sampling the full distributions of key model outputs. At present, the model is extremely restricted, so a MCMC is unlikely to be effective. However, as the model is made more flexible and samples from broad possible parameter ranges, MCMC will become very useful and better sample the uncertainty in the likelihood function (compared to jitters). MCMC jumps may also be considered, but again would be less necessary as the model becomes less restricted.

Recommendation 36. Undertake MCMC when the model parameter inputs are less restricted.

7.4.6 Summary

1. The CPT has produced a very rigorous set of guidelines (prepared by Punt and Kinzey) of which diagnostics and tests should be provided. Those provided are supported.

2. Extensive diagnostics are provided as a norm when considering the 2016 assessment report.

3. Undertaking MCMC is recommended, but only when the model is less restricted.

4. Jitter runs were not fully investigated. Results during the review show that these are important.

5. Profile likelihoods were attempted, but an ADMB bug made this impossible during the review. These should urgently be investigated given the model test results.

6. Different additional retrospective tests are suggested.

7.5 ToR 5: Recommendations for integrating BSFRF surveys into the assessment.

Side by side experiments between a BSFRF funded vessel using nephrops gear and the NMFS surveys have been performed. These experiments, aimed at tanner crab, have been undertaken in 2016 and 2017 (and planned for 2018). The nephrops gear is smaller but heavier. As a result, the gear samples a wider size range of crabs and also samples animals buried in the substrate. Both gear types use net geography, so swept area is calculated accurately. The nephrops trawl always pairs to the beginning of the NMFS 30 min site, but runs for 10 mins only. This may be causing some bias as the NMFS survey selectivity changes over the extent of the trawl. For this reason, it is more appropriate that either more 5 minute samples are taken during the 30 minute survey (probably not practical), or that the start location of the nephrops trawl is more randomised.

Recommendation 37. Investigate the possibility of increasing the number of samples within a site or randomizing the start location of the nephrops trawl relative to the NMFS survey.

Red king crab data comparing the CPUE ratio between the NMFS and the BSFRF vessels do show that the selectivity, but also the catchability can be different between years. For example, the 2014 year ratio is much higher than other years. This leads to three important conclusions: that the annual variability in NMFS survey catchability and selectivity may be far greater than previously thought, that the side-by-side series would be of greater benefit if continued in the medium term, and that the design needs to be consistent enough to ensure that it is useful for the assessment. It is therefore essential that some on-going sampling be considered.

Recommendation 38. The benefit of the side-by-side experiment, although short, has shown great value in understanding NMFS survey catchability and selectivity. However, it is highlighting strong inter-annual variation which means that the on-going sampling in the medium term should be considered.

An important aspect with regard to the utility of these data, is how to include them into the assessment. Importantly, they should *not* be treated as another index of abundance. These data are not independent of the NMFS survey data and inclusion as another independent index of abundance would imply a certainty that is not justified, i.e. there is an element of double counting.

The proposal provided for review does not include this option and indeed directly links the two assessments. The suggested approach is to assume that the nephrops catchability is 1 for all sizes included in the assessment. This assumes that the size composition of the nephrops gear is *exactly* reflecting the size-specific availability without noise. This is a crucial assumption which is not likely to be correct given there are occasions when the nephrops gear catches no crabs while the other does and *vice versa*. This assumption is even harder to justify when considering that the extent of the side-by-side experiments at this stage cover a smaller area than that undertaken by the whole NMFS survey. The fact that the 2018 survey is intended to include the northwest of this EBS is of great benefit.

Despite the above concerns, these data are extremely useful and can be included in the assessment in several ways.

The first is as suggested by the assessor, which is to include it through its size frequency and assuming that its catchability is 1. The equations then follow as described in the paper BSFRF Side-by-Side Results.

The second is to include the annually varying error in the assumption of unity by carrying over an error term associated with the $A_{y,z}$ into the equation calculating the $n_{y,z}^{NMFS}$ or directly estimating $A_{y,z}$ .as in the snow crab assessment. This error term would draw on the inter-site variability of *A* between the side-by-side trawls within a year from the provided data. Given there is a different extent of the survey per year, this error term may need to be different annually depending on the inter-annual variability obtained from the analyses. Although this second approach would reduce the value of these surveys to the assessment, it would be more consistent with the data in that error in availability would occur for both gear types, but less so for the nephrops gear. These two approaches would include the raw nephrops size frequency data and allow the data to influence the results. The former approach includes more error than the latter.

The third approach would be to place the ratio data into the assessment as shown in the presentation provided. The error in this ratio could then be included. Although this is ultimately mathematically similar to the above, a different data set is included (being the data). This ratio could then be estimated using a random effect in the model.

Recommendation 39. A staged approach is suggested for inclusion of these data, starting with the option provided by the assessor. The second would be to include the error term to investigate the importance of the fully sampled assumption. This latter is the preferred approach. A final option is the inclusion of the ratio data as a random effect, although these size data are relatively noisy.

One additional benefit of the shorter nephrops trawl is that some indication of the within grid/site variance can be established. These do not need to be undertaken as side-by-side trawls, but experiments independent of the NMFS surveys, other than being within the same grids and time scale. This would again give some indication on the variability around the nephrops availability and size frequency. This would also provide crucial data that would assist in the appropriate inclusion of the nephrops side-by-side surveys.

Recommendation 40. Consider an option to undertake independent BSFRF surveys that investigate the within grid patchiness of crabs.

7.5.1 Summary

1. Including the side-by-side survey data in the assessment is supported.

2. The approach suggested by Buckhausen as a first cut is also supported.

3. Estimating with site variability for the nephrops surveys are recommended.

4. A few alternative approaches to the data inclusion are also suggested.

5. Continuing this data source is recommended, given the short-term utility of the data in terms of selectivity and also relative catchability.

7.6 ToR 6: Recommendations for alternative assessment/projection model configurations.

The model seems over-informed and has too many different optimization systems (informative bounds, priors, sample sizes, penalties and weights) that make it difficult to determine all their interactions and driving forces. The next steps would have to be heuristically undertaken, at least in the initial process. The reason is that the result of one step will tend to determine the result of the next step. It is only feasible then to describe a process and practical options, rather than a single set of mathematics. There are many different approaches that could be used, but the below is the common approach and used in some Alaskan crab assessments as well.

7.6.1 Bounds etc.

1. The highest priority would be to remove as many as possible, or all of the informative bounds and priors in the model that include (at least) changes such as the inclusion of the EBS tagging data, estimating the growth scale parameter, etc., as per the CPT suggestions. The inclusion of the EBS data would be important in providing some information on EBS growth, rather than only using informative bounds/priors from Kodiak parameters estimated externally to the model.

2. Remove any of the bounds or priors that were included to restrict a parameter within certain natural bounds, e.g. 0 and 1, and instead use transformations to produce the same effect.

3. Initially set the likelihood weights up such that their role is primarily to turn the influence of data on and off, or to undertake discrete sensitivity tests.

Recommendation 41. Transform rather than use bounds, priors or penalties, if at all possible.

Recommendation 42. Use a combination of a few bounds/priors, etc., on a component likelihood.

7.6.2 Sample sizes

4. Calculate the actual variances external to the model, but estimate the effective (rather than actual) sample size for the size data in stages as described in McAllister and Ianelli (1997) and in more detail for crab in Siddeek et al. (2017) internal to the model. Although there are several approaches that could be applied, applying similar approaches as a default between the crab fisheries is recommended.

An alternative approach to calculating the effective sample size for size data is that by Folmer and Pennington (2000). A Dirichlet-multinomial distribution can be used to fit these records using a likelihood maximization technique given by Minka (2012). This approach is demonstrated in Deng et al. (2015).

Recommendation 43. Use iteratively, an approach that estimates effective sample sizes where appropriate, rather than likelihood weights (beyond Boolean weights). Ideally, use the approaches described in Siddeek et al. (2017) as a first step in an attempt to keep the Alaskan crab models relatively simple. This should be undertaken in combination with other changes, given these approaches do not fix model structure issues.

7.6.3 Data conflicts

8. Investigate the conflicts between the data and inputs to understand which model is most influential. Is the problem with this model caused by mis-specification or conflicts in the data? The role of each data set can be included and excluded. The role of different model specifications can be determined through an iterative approach of changing settings/assumptions within the model. If there is a need, a statistical design can be undertaken as per Dichmont et al. (2006a). This tested model sensitivity, and although initially an unbalanced design was produced, analyzing key results and then taking a smaller sub-set of the tests that were shown to be statistically significant resulted in a balanced design.

Recommendation 44. Undertake a series of tests removing data. Consider undertaking this using a statistical design approach.

The author is not in favour of blind model averaging techniques for integrated size based models, mainly because these models are often very sensitive, and not all model versions are equal which often means some *ad hoc* model weighting approach is needed. Also, the current model seems to have structural issues and data conflicts that should first be addressed.

See ToR 7 regarding simulation tests, as this interacts with model specifications, etc. **<u>Before moving to a major restructure of the model, the simulation tests may be a better first step</u>.**

7.6.4 Summary

1. The model is too restricted, yet there are clear conflicts with the data given the correlations and bounds being hit.

2. A series of steps to move forward are suggested. These include removing any bounds or priors that could be solved through transformation, calculating the effective sample size and using a sampling design to test the impact of different sample sets.

3. However, major structural changes are not recommended without undertaking a simulation study, which is highly recommended. In-depth suggestions regarding model structure are only really feasible when the current model and its variants are tested using simulated data.

7.7 ToR 7: Recommendations for research that would reduce the uncertainty associated with key parameters assumed or estimated in the assessment.

7.7.1 Growth

Extensive research is underway to collect the data needed to estimate EBS growth within the model from EBS tagging data. This is clearly a priority given that the model is sensitive to the Kodiak priors or inclusion of the EBS tagging data. Although the Munk data apply to the Kodiak region, they are undertaken over a longer period of time, and it would be worth investigating whether these data could provide insight into potential inter-annual changes in growth-per-molt in terms of estimating the effects of temperature, for example, on growth.

Recommendation 45. Support the on-going focus on estimating the EBS tanner crab growth per molt work and see this is a priority.

7.7.2 Skip molting and maturity

There was some discussion on whether there could be skip molting in tanner crabs or whether this is a case of the timing of the molt relative to the survey. There was also discussion of whether immature old shelled animals are classified correctly. Maturity is allocated based on chela height relative to carapace width. All animals above a specific ratio are deemed mature. Based on information provided in the workshop, redefining immature old shell as immature new shell affects the probability of the maturity ogive. Although this is a preliminary analysis, it highlights that these set assumptions regarding molting and maturity above a specific ratio may have some error associated with them and should be reviewed. This is important, given that an animal that matures at a small size in the model (although there is a low probability of this happening) will not grow to capture size.

Recommendation 46. Review the data and definition of maturity and its associated potential sources of error.

7.7.3 Space – east versus west

There appears to be spatial variation in much of the biology of tanner crab. This was identified with regard to maturity by Somerton (1981) and remains a feature of the NMFS survey data. There are indications of a north-west gradation with more mature males as survey sites move closer to the Pribilof Islands (PI). More mature females are found in the surveys in the offshore or the continental shelf sites. During the review, tagging to determine possible east-west movement was discussed and it is understood that this is hard to undertake as tags are mostly lost during molting. However, the use of transponders and pop-up tags is being investigated. This on-going work is supported given the importance of understanding more about the east-west movement of crabs, and also understanding the impact of recent spatial closures.

Survey data do not show that crabs move consistently between regions, but that this phenomenon is much more sporadic. Although there does seem to be some differentiation between east and west stocks, there may not be enough separation to produce genetically separate and independent stocks.

Similarly, the fleet is likely to change over time, given the above distribution changes and management actions such as spatial closures and the introduction of quotas.

Despite all these known spatial changes, the data analyses and the model essentially assume that these regions are similar or at least constant relative to each other over time. Given the difficulties with obtaining consistent and unique results for the model, at this stage this approach is appropriate. However, biological and fleet dynamic analyses may be an important first step using model based approaches, such as generalized additive models (e.g. Venables and Dichmont, 2004). However, as discussed in the simulation study below, some research steps towards understanding this gradation may be possible and feed into the necessary settings for the simulation study itself.

Recommendation 47. Undertake a detailed investigation of the fleet and biological dynamics over time, taking in consideration the requirements of a simulation study.

7.7.4 Role of predation and oceanographical signals on tanner crab biology and recruitment

The major predators of tanner crab include species such as Pacific cod and halibut, large tanner crabs, tom cod, and octopus. Most of these species are data rich, which would mean that some associated ecosystem work could be undertaken in the EBS. There are decadal scale oscillations in abundance that are reflected in earlier recruitment pulses. The fishery has experienced several periods of closures, either based on low female biomass or due to overfishing. The fishery is essentially based on periods of recruitment pulses as has been shown in several other fisheries, including for lobsters. The work of Livingston (1989) estimated predation mortality on 1 year old crab for 1981, 1984, and 1985, and these were high. Cod survey and assessment estimated biomass remained high for a longer period than the time block used in the model. This does beg the question as to why the enhanced M is for a shorter period than the cod data would imply. It points to the EM period being much more a construct of a need driven by the model than fully investigated by the theory.

Given the high value of fisheries and resources in this area and the richness of the data, an ecosystem model such as Atlantis (http://atlantis.cmar.csiro.au/) would be of high value. The operating model for this can be as complex or "simple" as required given the questions asked. For example, the Atlantis model can include as its focus all the fisheries in the region, focus on only one or on all the users. It is also a powerful tool to test across multiple species, for example, the crab tier system or the influence of climate change on the fisheries and associated harvest strategies. It is also spatially separated. Although it is a highly over-specified model, if used correctly in a relative sense (i.e. not as an assessment model) it is a powerful addition to the suite of single species assessment models available to the region, and can start addressing the more medium or intractable issues such as predation and climate change that could ultimately affect the assessment and harvest strategies.

Recommendation 48. Investigate the use of a full end-to-end model to investigate key issues such as predation impacts and the environment.

Apart from biological trophic relationship research, the wealth of data in the EBS would allow for oceanographic studies to address some key issues, such as the impact of the changes in the oceanography on crab larval distribution. There are often large changes to the bottom temperature with anecdotal information that there may be some effect on recruitment, although this relationship is unlikely to be simple. The larval and post larval stages are pelagic, which means that tools such as CONNIE2 (http://www.csiro.au/connie2/) and associated larval behavioral studies can be undertaken to test how the distribution changes, based on oceanography and other factors. CONNIE is one of several tools to model the movement of particles given the oceanography (the online version is a visualization tool and not the actual model). Handling mortality work also shows that tanner crab handling mortality in colder temperature periods may be higher than the 32.1% used in models.

7.7.5 Economics

The quota system in most fisheries has resulted in profound changes to the fleet dynamics of fisheries elsewhere. Since commercial catch rate is not included as an index of abundance, this is less of an issue in this assessment. Discussions at the workshop did not highlight that there had been large spatial changes to fleets' operations as observed elsewhere. If anything, it was pointed out that closures may have had a greater influence. Fuel prices were stated as being less influential for tanner crabs. Soak time, apart from a single outlier just after rationalization was similar, albeit slightly longer than before. Post rationalization, the pot limit was removed. Given the east-west gradients shown in tanner crab maturity, size, etc., the spatial dynamics of the fleet may become more influenced by external conditions (beyond those already identified). Several other fisheries of hard to age species have found recent profound changes to the selectivity of the fleet based on market demands. Discussion of this possibility pointed to this being less of an issue in this fishery. Despite this, it would be worth keeping a watching brief on this issue. One approach would be to investigate the difference in the size assumed discarded by the on-board observers and that measured by the dockside observers. There is no overlap between vessels sampled at sea and dockside by design. Despite this, an aggregated size frequency and shell state information would be useful. A further option would be to conduct some overlapping sampling of on board observer vessels and dockside.

Investigate whether there are differences over time between the dockside and on-board discard size frequency and consider overlap between on board and dockside sampling of discard data.

7.7.6 Bitter crab syndrome

Existing work on bitter crab syndrome is an important component of medium term research, although difficult to include in the model at this stage. On-going work in this field is supported.

7.7.7 Alternative models

Integrated size models are by nature very complex. The tanner crab model is no exception, especially since there are several fisheries that impact on tanner crabs. The model itself is also showing signs of data conflict, which are discussed above. However, these findings show that alternative models often provide insights that assist in the further development of the main model. Some of the key issues that cannot be addressed easily in the size model, are the role of the size data and the spatial changes over time. One of the types of assessment models that are very good at handling spatial data that are fairly subtle and not clearly stock based are hierarchical models. One example is a Bayesian hierarchical model applied to Northern Prawn Fishery prawns (Zhou et al., 2009). Although the formulation for the biomass dynamic component of the model is altered because of the short-lived nature of the prawn species, this aspect is easily changed. This reference is useful (even though there are many others that could be mentioned and the assessor should review the extensive literature on hierarchical models) since the model was tested against a delay difference model (Zhou et al., 2009) and later used within a Management Strategy Evaluation (Dichmont et al., 2008). Despite a standard biomass dynamic model not performing well for this species (Dichmont et al., 2006a), the hierarchical model performs as well as the delay difference and size model, indeed is used as a default for one of the species where the biology is less well known. The decisions as to which parameter are treated as hyper-priors in space would benefit the case here. Since only abundance data would be used, these could also inform the role of the size data in the model.

Recommendation 49. Develop an alternative model that can include spatial components, such as a hierarchical model (HIGH PRIORITY).

However, some of the spatial changes are size related and the hierarchical model would not be able to address these aspects. For this reason, model based size frequency distributions and survey abundance, amongst others, would provide great insight into the data themselves and confounding within the model. Covariates can then be added that may not easily be included in the model, such as bottom temperature. Although ideally the assumptions one would make should be internal to the model, e.g. using a Dirichlet or multinomial likelihood distribution, often developing model based aggregation approaches to the data outside of the model has great value. A large range of possible approaches could be undertaken. The simplest for the biomass would be to use a generalized additive model or mixed model. Given the survey covers such a broad range of species, the large range of species assemblage models would also apply if the work is expanded beyond tanner crabs. These approaches move beyond GAMs and GLMMs to random forests model (e.g. Wei et al., 2010; Pitcher et al., 2012).

Recommendation 50. Develop a spatial model of survey biomass and size frequency that includes habitat and covariates. Expanding this work beyond tanner crabs (HIGH PRIORITY).

7.7.8 Summary

1. The present research underway on growth, movement and bitter crab syndrome is supported as very important in the short and medium term.

2. A first step of undertaking a model approach (external to the assessment) of the spatio-temporal changes in the biology and fleet of tanner crabs should be done.

3. A medium term priority suggested is to undertake end-end whole of systems ecosystem modelling. The region is well studied with extensive data, and yet mortality (predation) and oceanography play an important role in the spatio-temporal distribution of the species.

4. Oceanographic modeling is also suggested, including ones that relate to larval distribution. An example of such a model is provided.

7.8 ToR 8: Suggested priorities for future improvements to the stock assessment/projection model

7.8.1 Simulation testing

Two major forms of simulation testing can be undertaken, Management Strategy Evaluation (e.g. Punt et. al., 2014) or simulation. MSE is most useful when trying to investigate various data analyses, model structures and assumptions to develop robust models. However, the nature of the process is that the assessment is not used to set the final TAC, rather stock status and the OFL and the ABC. As a result, simulation testing is perhaps the more reasonable approach. Here the simulation testing should investigate, for example:

1) The impact on the model of analyzing the data, considering the east-west maturity/size gradient on the data – this would require a more spatially complex model-based data analysis or changes to the assessment itself,
2) The sensitivity of the model to potentially different east-west population dynamics - this would require an east-west separation simulator,
3) The potential impacts of temperature and other environmental factors (see below) even if statistical relationships are weaker given the length of the data – this would require changes to the assessment such as recruitment covariances and correlates,
4) The utility of MCMC tests and whether perturbations are required to test the full chain,
5) The impact of different datasets on the model compared to the known simulator,
6) The impact of model structure, and
7) The impact of different likelihood formulations, e.g. Dirichlet versus multinomial and weighting techniques.

Recommendation 51. As the highest priority, undertake a simulation study to test various data analysis and model structure options, amongst others (HIGH PROIRITY).

7.8.2 Header file automation

Although not a priority, there could be much utility in automating some components of the model input file. This can be based on reading a set up file and then creating a default setup that reduces the chance of copying errors from one set up to the other. This can be undertaken in R, for example.

Recommendation 52. Create an automated input file generation system.

7.8.3 Summary

A major recommendation is to undertake simulation testing of the model. A series of points are provided that highlight some of the options that should be tested. This is the highest priority recommendation.

8 CONCLUSIONS AND RECOMMENDATIONS IN ACCORDANCE WITH THE TORS.

Summary of recommendations are provided below (those with high priority are noted):

9 REVIEW PROCESS

The review was undertaken in a very positive and helpful light. It was, however, noted that the review was undertaken mid-assessment cycle. This means that there was some confusion as to what assessment version was being reviewed. This was a bit uncomfortable in terms of documentation detail. There was also some overlap between the ToR. None of these issues seriously affected the review process, especially since the assessor was incredibly open to undertaking different tests and providing more detail.

10 REFERENCES

Zhou S, Punt AE, Deng R, Dichmont CM, Ye Y, Bishop J. 2009a. Modified hierarchical Bayesian biomass dynamics model for assessment of short-lived invertebrates: a comparison for tropical tiger prawns. Mar. Freshw. Res. 60(12): 1298-1308.

Tamone, S.L., Taggart, S.J., Andrews, A.G., Mondragon, J., and Nielsen, J.K., 2007, Ecdysteroid levels in Glacier Bay Tanner crab: Evidence for a terminal molt, in Piatt, J.F., and Gende, S.M., eds., Proceedings of the Fourth Glacier Bay Science Symposium, October 26–28, 2004: U.S. Geological Survey Scientific Investigations Report 2007-5047, p. 93- 96.

Deng, R. A., Punt, A.E., Dichmont, C.M., Buckworth, R.C., Burridge, C. Y. 2015. Improving catch prediction for tiger prawns in the Australian northern prawn fishery. ICES Journal of Marine Science. 72(1): 117-129.

Siddeek, M. S. M., et al. (2017). "Effect of data weighting on the mature male biomass estimate for Alaskan golden king crab." Fisheries Research 192: 103-113.

McAllister, M.K., Ianelli, J.N., 1997. Bayesian stock assessment using catch-age data and the sampling/importance resampling algorithm. Can. J. Fish. Aquat. Sci. 54, 284–300.

Folmer, O., and Pennington, M. 2000. A statistical evaluation of the design and precision of the shrimp trawl survey off West Greenland. Fisheries Research, 49: 165–178.

Minka, T. P. 2012. Estimating a Dirichlet distribution. http://research.microsoft.com/en-us/um/people/minka/papers/dirichlet/minkadirichlet.pdf.

Dichmont, C. M., et al. 2006a. "Management strategies for short lived species: The case of Australia's Northern Prawn Fishery: 2. Choosing appropriate management strategies using input controls." Fisheries Research 82(1-3): 221-234.

Dichmont, C. M., et al. 2006b. "Management strategies for short lived species: The case of Australia's Northern Prawn Fishery: 3. Factors affecting management and estimation performance." Fisheries Research 82(1-3): 235-245.

Venables, W. N. and C. M. Dichmont 2004. "GLMs, GAMs and GLMMs: an overview of theory for applications in fisheries research." Fisheries Research 70(2-3): 319–337.

Wei CL, Rowe GT, Escobar-Briones E, Boetius A, Soltwedel T, et al. (2010) Global Patterns and Predictions of Seafloor Biomass Using Random Forests. PLOS ONE 5(12): e15323. https://doi.org/10.1371/journal.pone.0015323.

Roland Pitcher, C., Lawton, P., Ellis, N., Smith, S. J., Incze, L. S., Wei, C.-L., Greenlaw, M. E., Wolff, N. H., Sameoto, J. A. and Snelgrove, P. V. R. (2012), Exploring the role of environmental variables in shaping patterns of seabed biodiversity composition in regional-scale ecosystems. Journal of Applied Ecology, 49: 670–679. doi:10.1111/j.1365-2664.2012.02148.x.

Punt, A. E., Butterworth, D. S., de Moor, C. L., De Oliveira, J. A. A. and Haddon, M. (2016), Management strategy evaluation: best practices. Fish Fish, 17: 303–334. doi:10.1111/faf.12104.

11 APPENDIX 1: BIBLIOGRAPHY OF MATERIALS PROVIDED FOR REVIEW

11.1 REVIEW PROCEDURE

1. Statement of Work: Bering Sea Tanner Crab Stock Assessment Review. National Oceanic and Atmospheric Administration (NOAA). National Marine Fisheries Service (NMFS). Center for Independent Experts (CIE) Program. External Independent Peer Review. Bering Sea Tanner Crab Stock Assessment Review. [SoW peer review_TannerCrab.pdf]

2. Tentative Agenda (2017-06-27). Bering Sea Tanner Crab Stock Assessment Review. [TannerCrab_CIEReviewAgenda.20170627.pdf]

3. Office of Management and Budget. Memorandum M-05-03. Final Information Quality Bulletin for Peer Review. [OMB_Peer_Review_Bulletin_m05-03.pdf]

11.2 DOCUMENTS (TO BE REVIEWED ARE IN BOLD)

11.2.1 Assessment documents

1. **Report to May 2017 CPT Meeting [201705ReportToCPT folder]**
    a. Stockhausen, W. T. (2017). Tanner Crab Assessment Report for the May 2017 CPT Meeting. Alaska Fisheries Science Center, April 2017.
    b. Appendix A: Corrected Retained Catch Size Frequencies in the Directed Tanner crab Fisheries.
    c. Appendix B: Tanner crab growth (molt increment) data.
    d. Appendix C: Tanner Crab Bycatch in the Groundfish Fisheries.
    e. Appendix D1: Model Comparisons for TCSAM2013 Models B0, B1, B2, B3, B4, B5, B6.
    f. Appendix D2: Model Differences for B1, B2, B3, B4, B5 and B6 vs. B0.
    g. Appendix E: Model Differences between T13B6 and T02A.
    h. Appendix F1a: Model Comparisons for T02A vs AG0.
    i. Appendix F1b: Model Differences for T02A vs AG0.
    j. Appendix F2a: Model Comparisons for AG1 vs AG0.
    k. Appendix F2b: Model Differences for AG1 vs AG0.
    l. Appendix F3a: Model Comparisons for AG1, AG2a, AG2b and AG3.
    m. Appendix F3b: Model Differences for AG1, AG2a, AG2b, and AG3.
    n. Appendix G1: Model Comparisons for TCSAM02 Models AG1, AG1a, AG1b and AG1d.
    o. Appendix G2: Model Differences for TCSAM02 Models AG1a, AG1b, and AG1d vs. AG1.
    p. Appendix H1: Model Comparisons for TCSAM02 Models AG1 and AG1c.
    q. Appendix H2: Model Differences for TCAM02 Models AG1c vs AG1.
    r. Appendix I1: Model Comparisons for TCSAM02 Models AG1 and AG1e.
    s. Appendix J2: Model Differences for TCSAM02 Models AG1 vs AG1e
    t. Appendix J1: Model Comparisons for TCAM02 Models AG3, AG3a, AG3b, and AG4.
    u. Appendix J2: Model Differences for AG3a, AG3b and AG4 vs. AG3.
    v. Appendix K1: Model Comparisons for TCSAM02 Models B1, AG4, and AG1c.
    w. Appendix K2: Model Differences for TCSAM02 Models B1, AG4, and AG1c.
2. **NPFSC BSAI Crab SAFE. (2016). 2016 Stock Assessment and Fishery Evaluation Report for the King and Tanner Crab Fisheries in the Bering Sea and Aleutian Islands [2016 Introduction Chapter Crab SAFE.pdf]**

3. **North Pacific Fishery Management Council's Crab Plan Team. (2016). Crab Plan Team Report. October 2016. [CrabPlanTeamReport_2016-10.pdf]**
4. **Punt, A., and Kinzey, D. (prep. by) Report of the Alaska Crab Stock Assessment Workshop, 13-14 May 2009. [SAFE_Guidelines.pdf]**
5. **Stockhausen, W. T. (2016).2016 Stock Assessment and Fishery Evaluation Report for the Tanner Crab Fisheries of the Bering Sea and Aleutian Islands Regions. September 2016 Plan Team draft. [TannerCrab201609_SAFEReport.pdf]**
6. Stockhausen, W. T. (2017). 2017 Crab Modeling Workshop Report for Tanner Crab. [TannerCrab201701_ModelingWorkshop.pdf]

11.2.2 Assessment model description

1. Anon. (no date) TCSAM02: The Tanner Crab Stock Assessment Model, version 2. [TCSAM02Description.201707.CIEReview.pdf]

11.2.3 Biology

2. Donaldson, W. E., Cooney, R. T., and Hilsinger, J. R. (1981). Growth, age and size at maturity of Tanner crab, Chionoecetes bairdi M. J. Rathbun, in the northern Gulf of Alaska (Decapoda, Brachyura). Crustaceana, 40: 286–302. [DonaldsonEtAl1981_Crustaceana.pdf]
3. Jadamec, L. S., Donaldson, W. E., and Cullenberg, P. (1999). Biological Field Techniques for Chionoecetes Crabs. University of Alaska Sea Grant College Program; Fairbanks, Alaska. [AK-SG-99-02PDF-Biological Field Techniqu.pdf]
4. Livingston, P. A. (1989). Interannual trends in Pacific Cod, Gadus macrocephalus, predation on three commercially important crab species in the eastern Bering Sea. Fishery Bulletin, 87: 807–827. [Livingston1989_FishBull.PredationMortality.pdf]
5. Paul, A. J., and Paul, J. M. (1995). Molting of functionally mature male Chionoecetes bairdi rathbun (decapoda: majidae) and changes in carapace and chela measurements. Journal of Crustacean Biology, 15: 686–692. [Paul_Paul1995b_JCrustBiol.Growth.pdf]
6. Siddeek, M. S. M., Zheng, J., Morado, J. F., Kruse, G. H., and Bechtol, W. B. (2010). Effect of bitter crab disease on rebuilding in Alaska Tanner crab stocks. ICES Journal of Marine Science, 67: 2027–2032. [Sideek2010_ICESJMS.BitterCrabDiseaseMortalityForTannerCrab.pdf]
7. Somerton, D. A. (1981). Regional variation in the size of maturity of two species of tanner crab (Chionoecetes bairdi and C. opilio) in the eastern Bering Sea, and its use in defining management subareas. Can. J. Fish. Aquat. Sci., 38: 163- 174. [Somerton1981_CJFAS.Maturity.pdf]
8. Stoner, A. W., Rose, C. S., Munk, J. E., Hammond, C., and Davis, M. W. (2008). An assessment of discard mortality for two Alaskan crab species, Tanner crab (Chionoecetes bairdi) and snow crab (C. opilio) based on reflex impairment. Fish. Bull., 106: 337–347. [StonerEtAl_FishBull2008_RAMP.HandlingMortality.pdf]
9. Urban, D., and Hart, D. 1999 Biology of the tanner crab Chionoecetes bairdi in Alaska: a report to the Alaska Board of Fisheries. Kodiak, Alaska: Alaska Dept. of Fish and Game, Division of Commercial Fisheries. Regional Information Report No. 4K99-22. [Urban_Hart_RIR.4K.1999.22.pdf]
10. Zheng, J. (2008). Temporal changes in size at maturity and their implications for fisheries management for eastern Bering Sea Tanner crab. J. Northw. Alt. Fish. Sci., 41: 137–149. [Zheng2008_JNorthwAtlFishSci.Maturity.pdf]

11. Zheng, J., and Kruse, G. H. (1998). Stock–recruitment relationships for Bristol Bay Tanner crab. *Alaska Fish. Res. Bull.*, **5**: 116–130. [ZhengAndKruse1998_AFRB.StockRecruitRelationships.pdf]

11.2.4 Fisheries

1. *Leon, J. M., Shaishnikoff, J., Nichols, E., and Westphal, M. (2017). Annual Management Report for Shellfish Fisheries in the Bering Sea–Aleutian Islands Management Area, 2015/16. Alaska Department of Fish and Game. Fishery Management Report No. 17-10. [ADFG.FisheryManagementReport17-10.pdf]*
2. *Alaska Department of Fish and Game. (no date). 2015–2017 King and Tanner Crab Commercial Fishing Regulations. Alaska Department of Fish and Game, Anchorage, Alaska. [ADFG.Regulations_KingTannerCrab-2015-2017.pdf]*
3. *Stockhausen, W. T. (2017). A brief history of the Tanner crab fishery. Unpublished. [TannerCr*abFisheryHistory.pdf]

11.2.5 Management

1. *Staff of the Alaska Fisheries Science Center, Seattle and Kodiak. (2008). Final Environmental Assessment for Amendment 24 to the Fishery Management Plan for Bering Sea/Aleutian Islands King and Tanner crabs to revise overfishing definitions. [NPFMC.FinalEAforAmendment24.pdf]*
2. *Anon. (2011). Fishery Management Plan for Bering Sea/Aleutian Islands King and Tanner Crabs. North Pacific Fishery Management Council, Anchorage, Alaska. [NPFMC.FMP_*Crab2011.pdf]

11.2.6 Surveys

1. *NMFS*
   a. **Daly, B. J., Armistead, C. E., and Foy, R. J. (2016). The 2016 Eastern Bering Sea Continental Shelf Bottom Trawl Survey: Results for Commercial Crab Species. NOAA Technical Memorandum NMFS-AFSC-327. [NOAA-TM-AFSC-327_2016_EBS_ShelfSurvey.pdf]**
   b. Somerton, D. A., and Otto, R. S. (1999). Net efficiency of a survey trawl for snow crab, *Chionoecetes opilio*, and Tanner crab, *C. bairdi*. Fish. Bull. **97**:617–625. [Somerton_Otto_FishBull1999.Selectivity.pdf]
2. *BSFRF*
   a. **Stockhausen, W. T. (2017). BSFRF Side-by-Side Survey Results. [BSFRFSurveys.pdf]**

11.3 ASSESSMENT MODEL RUNS [ASSESSMENTMODELRUNS FOLDER]

Base model [BaseModel folder]

ModelRun folder: Folder containing input data and model output for base model.

Miscellaneous data files for alternative model runs, together with R code to prepare model output for use with R

Base model plus growth data [BaseModel+GrowthData folder]

ModelRun folder: Folder containing input data and model output for base model.

Miscellaneous data files for alternative model runs, together with R code to prepare model output for use with R

Final model [FinalModel folder]

ModelRun folder: Folder containing input data and model output for base model.

Miscellaneous data files for alternative model runs, together with R code to prepare model output for use with R

Model comparisons [ModelComparisons folder]

R Markdown code and output for model comparisons and differences for B0, B1, and B2.

R packages [R_Packages folder]

Compressed files containing R code for utilities to process model output.

Notes on model runs [NotesOnModelRuns.docx]

Files containing executable code, i.e. runTCSAM02.pin.bat, runTCSAM02.pin.sh, tcsam02.exe, and tcsam02.osx)

12 Appendix 2: A copy of the CIE Statement of Work

Statement of Work

National Oceanic and Atmospheric Administration (NOAA)
National Marine Fisheries Service (NMFS)
Center for Independent Experts (CIE) Program
External Independent Peer Review

Bering Sea Tanner Crab Stock Assessment Review

Background

The National Marine Fisheries Service (NMFS) is mandated by the Magnuson-Stevens Fishery Conservation and Management Act, Endangered Species Act, and Marine Mammal Protection Act to conserve, protect, and manage our nation's marine living resources based upon the best scientific information available (BSIA). NMFS science products, including scientific advice, are often controversial and may require timely scientific peer reviews that are strictly independent of all outside influences. A formal external process for independent expert reviews of the agency's scientific products and programs ensures their credibility. Therefore, external scientific peer reviews have been and continue to be essential to strengthening scientific quality assurance for fishery conservation and management actions.

Scientific peer review is defined as the organized review process where one or more qualified experts review scientific information to ensure quality and credibility. These expert(s) must conduct their peer review impartially, objectively, and without conflicts of interest. Each reviewer must also be independent from the development of the science, without influence from any position that the agency or constituent groups may have. Furthermore, the Office of Management and Budget (OMB), authorized by the Information Quality Act, requires all federal agencies to conduct peer reviews of highly influential and controversial science before dissemination, and that peer reviewers must be deemed qualified based on the OMB Peer Review Bulletin standards. (http://www.cio.noaa.gov/services_programs/pdfs/OMB_Peer_Review_Bulletin_m05-03.pdf). Further information on the CIE program may be obtained from www.ciereviews.org.

Scope

The Alaska Fisheries Science Center (AFSC) Resource Ecology and Fishery Management (REFM) Division requests an independent review of the stock assessment/projection model used to conduct the Bering Sea Tanner crab stock assessment. The model is a size-based integrated assessment model and has been under continuous development since being approved for use by the North Pacific Fisheries Management Council (NPFMC) in 2012. It is anticipated that the North Pacific Fisheries Management Council's Crab Plan Team (CPT) and Science and Statistical Committee (SSC) will approve a change in the TCSAM (Tanner Crab Stock Assessment) code used for the assessment from "TCSAM2013", the code used for the 2013-2016 assessments, to "TCSAM02", a new modeling framework that provides a much more flexible environment than TCSAM2013 for defining alternative models based on a set of model configuration files, as well as fitting new data types not incorporated in TCSAM2013: molt increment (growth) and male chela height (maturity) data. TCSAM02 also calculates the OFL and associated quantities directly within a model run, and thus retains full model uncertainty when using MCMC, whereas using TCSAM2013 the OFL is calculated in a separate projection model and incorporates uncertainty only in recruitment and end-year mature biomass. This review will encompass the TCSAM02 stock assessment/projection model structure and assumptions on which it is based, as well as the life history, fishery, and survey data incorporated in the model. It will also address alternatives for incorporating several industry-funded surveys into the assessment. The Terms of Reference (TorRs) for the requested peer review are described in more detail in Annex 2.

Requirements

Three (3) CIE reviewers shall have the necessary qualifications to complete an impartial and independent peer review in accordance with the tasks and ToRs (Annex 2) described in the Statement of Work (SoW) herein. The CIE reviewers shall have expertise in conducting stock assessments for fisheries management and be thoroughly familiar with various subject areas involved in stock assessment, including population dynamics, size-structured models, harvest strategies, survey methodology, and the AD Model Builder programming language to complete the tasks of the scientific peer-review described herein. Familiarity with invertebrate stock assessment, knowledge of crab life history and biology, and harvest strategy development is desirable.

Tasks for reviewers

 Review the following background materials and reports prior to the review meeting:

1. Stockhausen, W. 2017. May 2017 Tanner Crab Stock Assessment Activities Report. In prep.

[For review:

2. Stram, D. et al. 2016. Introduction Chapter. In: 2016 Stock Assessment and Fishery Evaluation Report for the King and Tanner Crab Fisheries in the Bering Sea and Aleutian Islands. North Pacific Fisheries Management Council, Anchorage, AK. http://npfmc.legistar.com/gateway.aspx?M=F&ID=2f46b828-51ca-4a45-95bb-cddae2ed8f1d.pdf. [Review the "Stock Status Definitions" and "Status Determination Criteria" for background on the NPFMC's crab stock status criteria and OFL determination]

3. Stockhausen, W. 2016. 2016 Stock Assessment and Fishery Evaluation Report for the Tanner Crab Fisheries of the Bering Sea and Aleutian Islands Regions. In: 2016 Stock Assessment and Fishery Evaluation Report for the King and Tanner Crab Fisheries in the Bering Sea and Aleutian Islands. North Pacific Fisheries Management Council, Anchorage, AK. http://npfmc.legistar.com/gateway.aspx?M=F&ID=0e48278f-160e-426b-972e-f4736e7c8726.pdf. *[The last stock assessment, based on the TCSAM2013 model code.]*

4. DALY, B. J., C. E. ARMISTEAD, and R. J. FOY. 2016. The 2016 eastern Bering Sea continental shelf bottom trawl survey: Results for commercial crab species, 167 p. NTIS No. PB2016-104795. [Report on the 2016 NMFS annual eastern Bering Sea shelf summer crab/groundfish trawl survey.]

5. A document (TBD) describing the Gmacs assessment framework.

6. A document (TBD) describing the BSFRF surveys

 Attend and participate in the panel review meeting:

o The meeting will consist of presentations by NOAA and other scientists, stock assessment authors and others to facilitate the review, to provide any additional information required by the reviewers, and to answer any questions from reviewers.

 After the review meeting, reviewers shall conduct an independent peer review in accordance with the requirements specified in this SOW, OMB guidelines, and TORs, in adherence with the required formatting and content guidelines; reviewers are not required to reach a consensus.

 Each reviewer may assist the Chair of the meeting with contributions to the summary report, if required by the TORs.

 Deliver their reports to the Government according to the specified milestone dates.

Foreign National Security Clearance

When reviewers participate during a panel review meeting at a government facility, the NMFS Project Contact is responsible for obtaining the Foreign National Security Clearance approval for reviewers who are non-US citizens. For this reason, the reviewers shall provide requested information (e.g., first and last name, contact information, gender, birth date, passport number, country of passport, travel dates, country of citizenship, country of current residence, and home country) to the NMFS Project Contact for the purpose of their security clearance, and this information shall be submitted at least 30 days before the peer review in accordance with the NOAA Deemed Export Technology Control Program NAO 207-12 regulations available at the Deemed Exports NAO website: http://deemedexports.noaa.gov/ and http://deemedexports.noaa.gov/compliance_access_control_procedures/noaa-foreign-national-registrationsystem.html. The contractor is required to use all appropriate methods to safeguard Personally Identifiable Information (PII).

Place of Performance

Each CIE reviewer shall participate in, and conduct an independent peer review during, the panel review meeting at the Alaska Fisheries Science Center (AFSC) in Seattle, Washington. Pre- and post-review performance shall be conducted at the contractor's facilities.

Period of Performance

The period of performance shall be from the time of award through XXXXX. Each reviewer's duties shall not exceed 14 days to complete all required tasks.

Schedule of Milestones and Deliverables: The contractor shall complete the tasks and deliverables in accordance with the following schedule.

Within two weeks of award	Contractor selects and confirms reviewers
Approximately 2 weeks later	Contractor provides the pre-review documents to the reviewers
Approximately 2 weeks later	Panel review meeting
Approximately 3 weeks later	Contractor receives draft reports
4 weeks later	Contractor submits final reports to the Government

Applicable Performance Standards

The acceptance of the contract deliverables shall be based on three performance standards:
(1) The reports shall be completed in accordance with the required formatting and content (2) The reports shall address each TOR as specified (3) The reports shall be delivered as specified in the schedule of milestones and deliverables.

Travel

All travel expenses shall be reimbursable in accordance with Federal Travel Regulations(http://www.gsa.gov/portal/content/104790). International travel is authorized for this contract. Travel is not toexceed $XXXX – Mark Chandler will fill this out.

Restricted or Limited Use of Data

The contractors may be required to sign and adhere to a non-disclosure agreement.

Annex 1: Peer Review Report Requirements

1. The report must be prefaced with an Executive Summary providing a concise summary of the findings and recommendations, and specify whether or not the science reviewed is the best scientific information available.
2. The report must contain a background section, description of the individual reviewers' roles in the reviewactivities, summary of findings for each TOR in which the weaknesses and strengths are described, andconclusions and recommendations in accordance with the TORs.
a. Reviewers must describe in their own words the review activities completed during the panel review meeting, including a brief summary of findings, of the science, conclusions, and recommendations.
b. Reviewers should discuss their independent views on each TOR even if these were consistent with those of other panelists, but especially where there were divergent views.
c. Reviewers should elaborate on any points raised in the summary report that they believe might require

further clarification.
d. Reviewers shall provide a critique of the NMFS review process, including suggestions for improvements of both process and products.
e. The report shall be a stand-alone document for others to understand the weaknesses and strengths of the science reviewed, regardless of whether or not they read the summary report. The report shall represent the peer review of each TOR, and shall not simply repeat the contents of the summary report.
3. The report shall include the following appendices:
Appendix 1: Bibliography of materials provided for review
Appendix 2: A copy of this Statement of Work
Appendix 3: Panel membership or other pertinent information from the panel review meeting.

Annex 2: Terms of Reference for the Peer Review

Bering Sea Tanner Crab Stock Assessment Review

The report generated by the consultant should include:

1. Statements assessing the strengths and weaknesses of the current Tanner crab stock *assessment* model with regard to population dynamics, fishery and survey components, likelihood components, and model evaluation.
2. Statements assessing the strengths and weaknesses of the current Tanner crab stock *projection* model, with regard to methodology.
3. A review of the fishery dependent and independent data inputs to the stock assessment with regard to quality of information and appropriateness to the assessment.
4. Recommendations for alternative approaches to evaluate model convergence and compare multiple models.
5. Recommendations for integrating BSFRF surveys into the assessment.
6. Recommendations for alternative assessment/projection model configurations.
7. Recommendations for research that would reduce the uncertainty associated with key parameters assumed or estimated in the assessment.
8. Suggested priorities for future improvements to the stock assessment/projection model.

Annex 3: Tentative Agenda

Bering Sea Tanner Crab Stock Assessment Review

NOAA Alaska Fisheries Science Center
7600 Sand Point Way NE
Seattle, WA 98115
07-10 August 2017
point of contact: William Stockhausen; william.stockhausen@noaa.gov; 206-526-4241
Dates/times are tentative and subject to change

Monday, Aug. 07
09:00 Welcome and Introductions
09:15 Role of chair and reviewers, terms of reference
09:30 Overview (fishery, catch levels, bycatch, surveys)
10:30 Biology (growth, natural mortality, maturity curves, mating, molting frequency)
12:00 Lunch
13:00 Survey methodology
14:30 Fishery history and current operation
15:30 Harvest control rules and overfishing definition
17:00 Evening break

Tuesday, Aug. 08
09:00 Stock assessment and projection model
12:00 Lunch

13:00 Stock assessment and projection model (continued)
17:00 Evening break
Wednesday, Aug. 09
9:00 Current research studies
growth, fecundity and egg production
BSFRF side-by-side surveys and other research
12:00 Lunch
1300 Strategies for integrating BSFRF surveys into assessment
14:00 Gmacs
17:00 Evening break
Thursday, Aug. 10
9:00 Reviewer discussions with assessment author.
Review of requested model runs if required.

13 APPENDIX 3: PANEL MEMBERSHIP OR OTHER PERTINENT INFORMATION FROM THE PANEL REVIEW MEETING.

13.1 PRESENT (NAME, AFFILIATION)

1. Martin Dorn, AFSC, Meeting chair
2. William Stockhausen, AFSC, Lead assessment author
3. Jack Turnock, AFSC
4. Anne Hollowed, AFSC
5. Jeff Napp, AFSC
6. Ben Daly, Alaska Department of Fish and Game, Kodiak
7. Scott Goodman, Bering Sea Fisheries Research Foundation
8. Gary Stauffer, Bering Sea Fisheries Research Foundation
9. Alathea Letaw, University of Washington

13.2 REMOTE

1. Robert Foy, AFSC Kodiak
2. Miranda Westphal, Alaska Department of Fish and Game, Dutch Harbor

13.3 CIE REVIEWERS

1. Cathy Dichmont, Cathy Dichmont Consulting, Australia
2. Anders Nielsen, Technical University of Denmark, Denmark
3. Norman Hall, Murdoch University, Western Australia

AFSC: Alaska Fisheries Science Center